

Electronic Theory Handbook

\$1.65*
(N.Z. \$2.00)

INSIDE COMPUTERS

GET READY FOR THE HOME COMPUTER REVOLUTION

UNDERSTANDING ELECTRONICS

FROM AMPS TO ZENERS

COMMUNICATIONS TECHNOLOGY

KEEP IN TOUCH WITH THE WORLD

PERSONAL COMPUTERS AND YOU
MICROCOMPUTER CONNECTION
INSIDE MICROPROCESSORS
SHORTWAVE BROADCASTS

BASIC
COMPUTING
LANGUAGE

USE WITH CONSOLE KEYBOARD
CX14002

3 MEMORY
MODULE

3 MEMORY
MODULE

3 MEMORY
MODULE

OPERATING
SYSTEM

16K-RAM

16K-RAM

16K-RAM

10K-RO

ELECTRONIC THEORY HANDBOOK

Contents

INSIDE COMPUTERS

Personal Computers and You	4
Inside Microprocessors	7
Computer Graphics	9
Computer Communications	12
Operating Systems	14
Bubble Memories	16
Floppy Disk Systems	19
Microcomputer Connection	23
Keyboard Interfacing	24

COMMUNICATIONS TECHNOLOGY

Radio Receivers	30
Radio Transmissions	32
Signal Propagation	37
AM/FM Receiver Alignment	39
Satellite Receivers	42
Directional Antennas	44
Shortwave Broadcasts	47

UNDERSTANDING ELECTRONICS

Amps, Ohms and Volts	48
Prefixes and Exponents	54
Capacitance Basics	57
SCR Principles	61
Cathode Ray Tubes	63

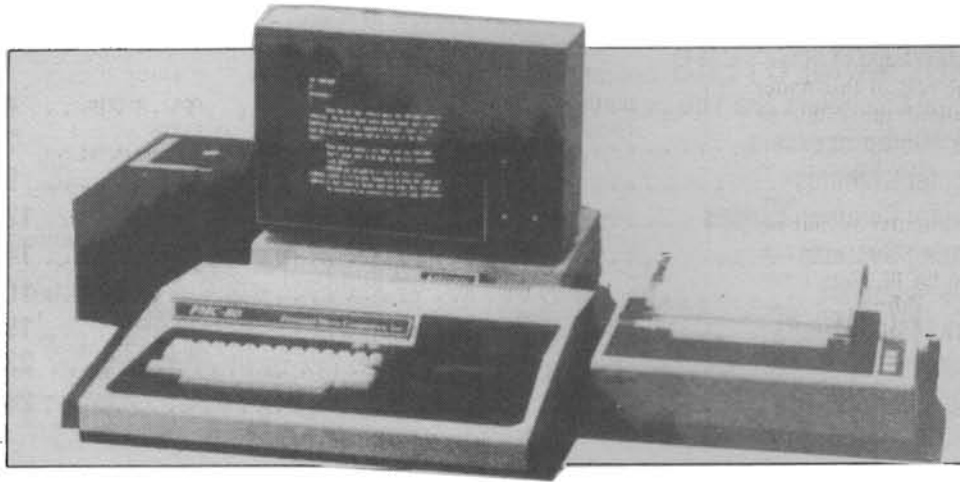
NOTE: The information published in this magazine originated from the U.S.A. Accordingly it is suggested that before undertaking any projects readers would be advised to check first the availability of the required parts.

Published by Page Publications Pty. Ltd. (Incorporated in N.S.W.) and printed by
Rotary Colorprint Co. Pty. Ltd. Both of Butt and Clisdell Streets, Surry Hills, N.S.W.
2010. Distributed by Gordon & Gotch Ltd., Melbourne-Sydney.

* Recommended and maximum price only.

Inside Computers

Personal Computers and You



□ If you don't know anything about home computers, you're nevertheless among the *majority* of people who will own and operate their own computers within the next two to three years. The fact is, more people will be buying computers for their own personal use (they're called "personal computers") in the next twelve months than own them already.

One reason for that is that computers are easier to use, easier to own, easier to expand and more capable than ever. Five years ago the guy with his own computer had to know the bits and bytes and boards inside his baby inside out—sometimes just to keep it running, or to get things back to normal after an all-too-frequent "crash," when his programming disappeared with a glitch on the power line or a bump on the chassis box.

Today, the housekeeping has been done by the companies that design these personal computers and related products. We see high-level programming languages like BASIC available

any time the power switch is on. By the way, when we say "high level," we're talking about the sophistication of the programming *inside* the computer that makes it simpler and easier for any computer user to work with.

The hard work has been done by the designers and manufacturers of today's more polished computers, peripherals, software, products and services. The fun lies ahead for you.

Your Faithful Ally. Your personal computer is a friend, an associate, a playmate, a servant and more. It'll act as a secretary, a typist, an accountant, a clerk, a teacher and more.

Want to go on an airplane ride from your favorite comfy chair? There are flight simulator programs that turn your computer into an imaginary airplane, giving you an animated pilot's-eye-view out the window (on the "TV" screen of your video monitor) that shows the swoops, dives, climbs and turns that you command through the keyboard. And this is only one of a family of programs called *interactive*

simulators that let your computer take you into worlds and activities you might never enter in "real life."

And there are interactive games that take you on all kinds of fantasy adventures, fighting trolls or pirates or outer space demons, changing the game with every move you make.

On a more practical level, there are programs like VisiCalc™ that perform like an electronic worksheet, doing calculations for you and letting the results react to changes you make in specific entries. You can enter descriptions, formulas or values in any of hundreds of specific places, get answers to "what if" questions, look for highs and lows, graph your results and more! It's like having a mathematician, an accountant and a business advisor all rolled into one. And it's just one of the things your own computer can do for you.

How about a secretary and typist—one who never makes a mistake—and lets you type perfectly, even when you make mistakes? That's what word

Inside Computers

processing programs do. You can personalize a "form" letter you've written for someone else; write, modify and reproduce a manuscript; let the computer retype something with different spacing or margins and do a score of other tricks that would take a lot of time and trouble if not for your friendly personal computer.

And we've hardly just begun. Take a look at some of these software categories (*software* or *programming* are the terms used to refer to the prepared lists of instructions a computer needs to follow in order to accomplish any given task; you can prepare them yourself by writing a program or you can buy them ready-to-run) to start getting a feeling for some of the things you can do with your own computer:

And that's only the beginning. As you read through the rest of this issue, you'll learn more about the different kinds of software and what they do.

Speak to Me, Write to Me. One of the nicest things about what you can do with your own computer is that no matter how small you start, you can usually keep on adding to your computer's capabilities by adding *peripherals* (the term for pieces of equipment specifically designed to work with your computer).

There are *printers*, priced from a few hundred dollars to several thousand, that take the work you do with your computer and give you a permanent record in black-and-white (and sometimes, colors).

And *plotters* that draw pictures as expertly as an artist or draftsman, thanks to the capabilities of your computer. In fact, some printers have a plotting capability included.

Your computer can even talk to you, thanks to a relatively new kind of peripheral called a *speech synthesizer*. These are getting better and better (and less and less expensive!), and giving computers the power of spoken language in communicating with you . . . all under the command and direction of your software.

On the other end, there are also *speech recognition* peripherals that let you train your computer to recognize a small vocabulary of commands that you speak into it (through a microphone).

You can also *draw* information in a way your computer can recognize with *digitizing pads* and *tablets*.

There are a number of peripherals called *power controllers* that let your computer turn on an off lights and

Entertainment

- board games
- space games
- simulations and other games

Graphics

- charting data
- creating pictures
- marquees and signage
- animation

Hobbies

- model aircraft design
- astronomy
- astrology and biorhythms
- solar energy
- satellite tracking
- electronic design

Leisure

- music makers
- poetry writers
- contests

Education

- math
- foreign languages
- speech
- history
- social studies
- drivers' ed
- geography

Home

- personal fitness
- nutrition
- personal finances

- electronic checkbook
- appliance controller
- alarm/security
- recipes
- meal/menu planning

Math

- statistics
- trigonometry
- conversions

Personal Business

- stock analysis
- income tax
- loans
- investments

Business

- general ledger
- accounts payable/receivable
- payroll
- inventory
- dispatching
- mailing lists
- real estate
- taxes

Files/Data Bases

- file management
- data retrieval
- data management

Words and Lists

- word processing
- letter writers
- mailing lists
- address lists

appliances around your home, plant or office. Imagine your computer (when it isn't otherwise occupied) waking you up, turning on the bedroom and bathroom lights, brewing your coffee and frying your eggs, all according to the schedule you've programmed in—and this only scratches the surface of the usefulness of these delightful peripherals.

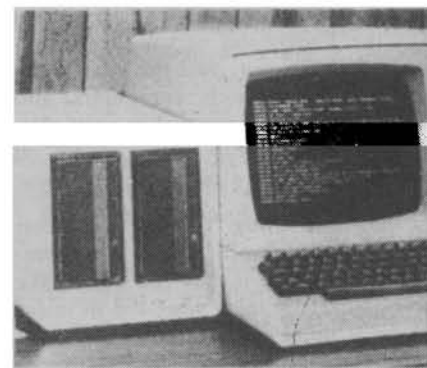
And for those of you with a yen for mechanical experimentation, there are even *robots* that attach to your computer like other peripherals and let you find new ways for your computer to do things for you—including helping you have fun.

Computer to Computer. One of the most exciting things you can do with a personal computer today is to let it talk to other computers, through a simple peripheral called a *Modem* (it stands for modulator-demodulator, which is one way of saying talker-listener) that connects your computer to your telephone line, and through established *networks* that let you and your computer use this telephone hookup to contact other computers—including some big, special ones.

You can join these networks for

very nominal fees (plus so much per minute of use, like your telephone when you call long distance). There are several already in service, and more planned all the time. Here's some of what you can expect to be able to do via today's and tomorrow's networks with your own personal computer.

You can tap into the press news wire service feeds—the teletype services that send news to local newspapers and broadcast newsrooms around the world—and read the news



This Heath setup has all the basics: keyboard, memory and video monitor. The computer itself is not shown here.

on your tv screen as it's sent down the wire (why wait for tomorrow morning's paper?).

You can keep on top of the stock market. Read a computerized version of classified ads, where people offer to buy sell or trade things with others on the network. There are message centers, bulletin boards, all kinds of channels.

There are also active communications going on. One network has set up kind of a computer citizen's band, with forty channels on which round-table "gabfests" (with typed messages via computer keyboards appearing on the screens of whomever's plugged into the network) are encouraged. And there are more private, one-to-one communication channels available to let you and your computer talk to someone else and his computer—all you need to know is his access address, the network equivalent to a phone number.

And more services are coming. In the near future, area businesses will be offering specific services, either established networks or through their own telephone hookup. Imagine being able to link to your bank by computer, typing in a secret access code and making payments, checking your balances, applying for a loan or a credit card and so on, all from the convenience of your home or office computer.

Another service might give your computer a listing of the week's tv programs, or schedules and reviews of movies in the area. Or you might be able to check on the prices of groceries at area stores. How about computer catalog shopping by credit card? The money you save on gas alone might help buy you that next program or peripheral!

And there are groups of computer users who network themselves together (as few as two at a time, as many as dozens) just to play games! It's the Federation versus the Romulans, the Klingons and the Renegades of Zarn—with both computers and thinking players involved in very realistic gamesmanship. Then again, it could just be a game of checkers.

Plug and Chug. One of the nicest things about owning your own computer these days is that you don't have to be an electronics whiz kid in order to set it up, use it, and get everything out of it that it has to offer.

It's as easy as plugging into the power outlet on the wall and turning the switch to "on." You may have to



One of the newest mass-market personal computers is the Atari 800. It has plug-in RAM and ROM cartridges available to increase versatility, make upgrading a user operation.

make a simple connection to your TV set, in some cases (and it'll still work as a TV set when you want it to). There are even small, portable computers—they're shrinking down to pocket size, and are almost as easy to use as a pocket calculator.

It used to be that the only way to enter a program that you wanted the computer to run was to type it in via the keyboard every time, or possibly try to save it on a cassette. In the early days, cassettes were "iffy" things, slow to work with, and since they weren't designed for computers, they were liable to lose bits of data.

Today, specially engineered data cassettes, improved cassette handling capabilities by computers and the easy, inexpensive availability of pre-



Having a personal computer in your own home opens up a world of programming fun for the whole family.

pared programs already on cassette (and quality-checked to assure they'll run) have taken a lot of the old "nervous" out of programming—even for computers on a tight budget.

But the tight budgets of even just five years ago will buy a pretty sophisticated system today. Today's computers cost as little as \$200-\$600 for tight-budget machines. For \$1000-\$2000, you can buy some marvelous extra capabilities.

Part of the difference is in the amount of RAM (memory storage capability) a computer has—the more RAM, the more complex a program it can handle, or the more data at once.

Another difference is in the way a program is loaded. Disk memory (generally in the form of a flexible magnetic "platter" in a protective cardboard sleeve) offers several advantages. For one, programs that would take hours to load by hand or minutes to load by cassette can be loaded (the term used when a program is played into the computer to give it the instructions it will follow) in seconds with a disk. For another, a disk can hold a much longer program than a cassette can. And for a third, a disk permits you to store data (specific pieces of information, like the words of a letter or the amounts and numbers of your checks) in addition to the programming that uses it—something cassettes *can* do but disks do better and easier. Disk drives (the peripheral equipment that records and plays a disk, or several disks) can cost as little as a few hundred dollars.

Another part of the difference between machines lies in the capabilities of their displays. Some include built-in LCD (liquid crystal) or LED (light emitting diode) rows of characters—like a calculator, but with more character positions and with the capability of forming letters and symbols in addition to numbers. Others hook up to a TV set or video monitor, with the capability of displaying only so many characters per line, and only so many lines per "page" (page in terms of a picture tube means at one viewing, in other words, only so many lines on the screen at one time).

Video graphics vary, too. Some computers use large blocks (called *pixels*, which stands for picture elements), which may be turned on or off or "moved" around the screen. The larger the block, the cruder the pictures the computer can portray with them. Also, the larger the block, the fewer of them can be fit onto the screen at once. Graphics that use large

Inside Computers

blocks are generally referred to as offering *low resolution graphics*.

On the other hand, some computers are designed to be able to use very small blocks, which provide very refined pictures. Curved lines look more natural, less like they have "sawtooth" or "staircase" edges. Straight and curved lines can be made much thinner, permitting much finer detailing. Since a very large number of these blocks can fit on a screen, this is referred to as *high resolution graphics*.

Other differences in graphics include black-and-white versus color capabilities, and the amount of resolution available in color versus the resolution available in black-and-white.

Another way different computers differ from each other is in the way you type in information or programming or respond to questions. Some computers offer tiny, calculator-size pushbuttons—hardly encouraging for a touch-typist. Others offer flat keyboards, where you press a drawing of a key (there's a thin "membrane" switch behind each drawing). Some are organized typewriter-style, others in alphabetical order (though this is

increasingly rare). Some look exactly like typewriter keyboards, and do everything except pound type against a ribbon. Some offer a combination of a typewriter keyboard and a separate set of keys that make entering numbers more convenient. Still others offer special functions. Some click when you press the keys, some beep or boop, others don't make a sound.

Your best bet is to try a couple of different keyboard arrangements to see which you're comfortable with.

Try It Yourself. You don't have to come in waving a roll of hundred dollar bills in order to get the attention of a salesman at your local computer store, all you need is an interest in computers. Most salesmen are happy to show you what today's personal computers can do. They'll help by asking questions about your interest in computers—is it for home? for work? for school? for fun? do you need color? a printer? disk memory? a modem? is it for now? this year? next year?—to help steer you to exactly the kind of system that best suits you.

Many computer stores offer free or inexpensive classes in computers. Some

even sponsor regular open house events or shows.

Remember, most of the people who buy computers this year know little or nothing about them, except that they have a lot to offer. Computer stores know this, and are anxious to help people understand how to choose the right computer for each individual's needs.

That's what we're doing, too. If you already have a computer, you'll find these pages chock full of information on equipment, peripherals, systems, software and more—plus, a few practical (and a few fun) programs you'll want to enter and run.

If you haven't bought yourself a computer yet—and, remember, that puts you in the *majority* of people who will own their own computers a year from now—you'll find the information in these pages even more interesting.

Today's computers are fun, fascinating, easy to use, inexpensive and just waiting for you. What can you do with one? More than you've ever imagined! ■

Inside Microprocessors

[1] Most home computers can be managed quite nicely by learning one or more of the popular high-level languages like BASIC or PASCAL. That does not, however, prevent us from being a trifle curious about what is really going on inside the machine. One can drive a car quite well without any idea of how the engine actually works, and one cannot repair a car without devoting more attention and study to its functioning than most of us can afford. Many people, however, get a degree of satisfaction from appreciating the basic concepts of how it works. The same can be said of computers and the people who use them.

It is a well-known fact that a digital computer consists of an input output device, central processor and memory, and that it does all its work by manipulating 1's and 0's which are represented by *on* or *off* conditions of electronic switches. But how is all this activity organized and how does the machine accomplish all its complex computations at such rapid speed.

Communication. To answer this question, we must take a quick look at how the central processing unit, or CPU, communicates with computer memory and then a more detailed look at what goes on inside the CPU. The simplest way to think of the CPU is as a mailman, who goes from one numbered mailbox to another in succession. The mailboxes are locations in computer memory which contain program instructions or data for the CPU. The numbers of the mailboxes are the *memory addresses* and it is to these addresses that the CPU refers as it steps through a program.

For instance, when the CPU looks at an address, it may find an instruction telling it to go to the next address and read the contents of that location into one of its *registers*. CPU registers are special memory locations inside the CPU that are used for temporarily storing and manipulating data. Because of its function, this type of memory is often referred to as "scratchpad memory."

Going to the third address in line, the CPU may find a similar instruction which tells it to take the contents of the fourth location and store them in another CPU register. At this point, the CPU has two different numbers stored in two of its internal registers. The next instruction the central processor sees may be a command to add these two numbers together, that is, to add the contents of one register to the contents of another. The final instruction might then tell the CPU to take this result and store it in another specified location out in computer memory, or to output it to an output port where it could be read by a terminal device.

Our simple example has shown that a computer program consists of two things: *Instructions* to the CPU to perform certain operations, and *data*, or information that the CPU needs to carry out these instructions. The latter may come in various forms and we will discuss only a few of them here.

Architecture. Looking into the CPU, that microscopic silicon wonder, we

find that it contains a number of distinct internal components. The ones we will discuss here are the registers, the program counter and the controller. The number, size and arrangement of these internal components is known as the CPU's *architecture*. For the purposes of this article, we will refer to the architecture of the popular 8080 8-bit microprocessor.

Since the length of a data word for the 8080 is 8 bits, or one byte, each internal register will be able to store one byte. Two exceptions to this are a couple of specialized registers called the stack pointer and the program counter. These registers must store memory address information, and since addresses in the 8080 can be up to 16 bits in length, these two registers must be able to store two bytes.

We will not discuss the stack pointer at this time, but the program counter will be of central importance. You will notice I listed the program counter as a CPU component in its own right because of its importance, but it is really a specialized register.

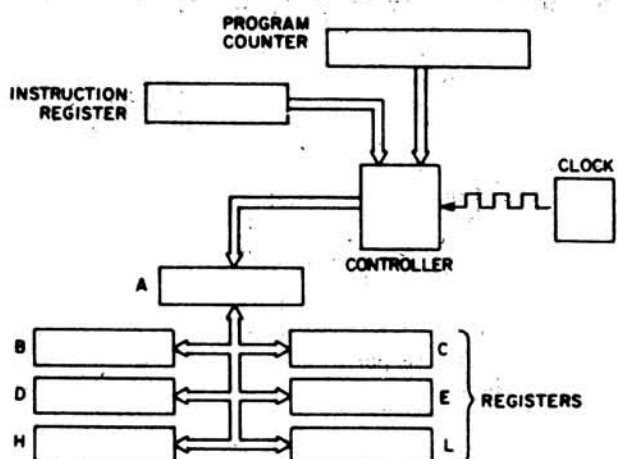
Let us imagine our simplified 8080 as illustrated in figure 1. We see the different registers with connections between some of them so that data can be easily transferred among them. Also illustrated is the controller with the clock input. Other CPU components include the arithmetic logic unit, the address buffer, and the instruction decoder.

What makes the whole thing go? That is the function of the system clock, which is usually connected externally to the CPU. It is a crystal oscillator which produces a rapid series of pulses which are used as a timing reference by the entire system. It might be more useful to think of it as a metronome rather than a clock, as some functions require a different number of clock pulses to be performed than others. Every operation of the machine, however, uses this reference to time the sequence of its operations. In an 8080-based system, the clock frequency is typically 2MHz—which may do something to explain the speed with which a computer computes.

Operation. Let us now perform an actual addition of two numbers while examining what is taking place inside the CPU. To do this, it will be necessary to have an addition program contained in the computer's memory. As mentioned above, such a program will consist of instructions and data. The instructions are binary codes that cause the CPU to perform certain operations on the data. There is a limited number of instructions any given CPU can perform, and the list of these instructions

Figure 1

The important internal components of the 8080 central processing unit (CPU) are shown in this block diagram. The data is transferred along the wide lines according to the instruction in the instruction register. The clock keeps it all in perfect timing.



is known as the *instruction set*.

A computer program is a list of appropriate instructions, also known as *opcodes*, and their *operands*, or data the CPU needs to carry out the instructions, stored in an orderly manner in memory.

To execute the program, the *program counter* goes to the address of each instruction in turn, loads that instruction into the instruction register, and then goes to the address of the next instruction. The instruction in the instruction register is decoded by the instruction decoder and executed and the process is repeated until a programmer's error is encountered and the machine gets confused.

When writing programs such as these in machine language, the individual CPU's architecture is a major consideration. The nature of the machine's instruction set is based upon its architecture, and in order to program for a different CPU, a different instruction set must be learned. High level languages like BASIC and PASCAL allow us to communicate in English-like statements of the problem. The interpreter or compiler translates the program into machine instructions the CPU can understand.

For our purposes, let us assume that our addition program has either been entered in machine language by a programmer or been translated by a compiler. At any rate, it now resides in memory as absolute machine code—the stuff our CPU can digest. Also, for our purposes, we will not use a terminal (input/output can really get complicated) and assume that the two numbers to be added are contained in the program.

Instruction. While there are perhaps more elegant and efficient ways to write this program, it will serve to illustrate some of the things the computer does.

To the CPU, our program would look like this:

```

1. 00100001
2. 10000000
3. 00100000
4. 00000110
5. 00000101
6. 00111110
7. 00000010
8. 10000000
9. 01110111
10. 01110110

```

This is, of course, incomprehensible to a human reader, especially since practical programs are many hundreds of times longer than this. For this reason, those bytes which are instructions and not data can be described by humans with short abbreviations called *mnemonics*. Mnemonics apply only to instructions in the instruction set and numeric data must be represented for humans by another form of notation than binary, usually octal (base 8) or hexadecimal (base 16). We won't discuss these here, however, and I will use decimal notation to indicate numbers the computer must deal with.

Remember that each 8-bit byte of code above resides in a separate location in memory and that each such location is specified by a unique address. For the 8080, addresses can be numbered from 0 to 65,535 decimal.

If we were to translate this program into readable form, we could see that it is telling the CPU to first load the H and L registers with the numbers in the next two memory locations. The first goes into the L register and the second into the H register (this reverse order is an 8080 convention and not true for all CPU's). This number will later be used by the computer to refer to an address elsewhere in memory.

Inside Computers

The next *instruction* (not the next byte) is at address number 4 in our program. It tells the computer to put the following number (here, the number 5) into the B register. The next instruction similarly tells it to put the contents of the next memory location into the A register. At this point in the program, the computer registers look like figure 2. Note that the program counter contains the address of the *next* instruction to be put into the instruction register.

That instruction is the ADD B instruction that tells the CPU to add the contents of register B to the contents of register A. The result will be the new contents of register A, but the data in register B will not be destroyed.

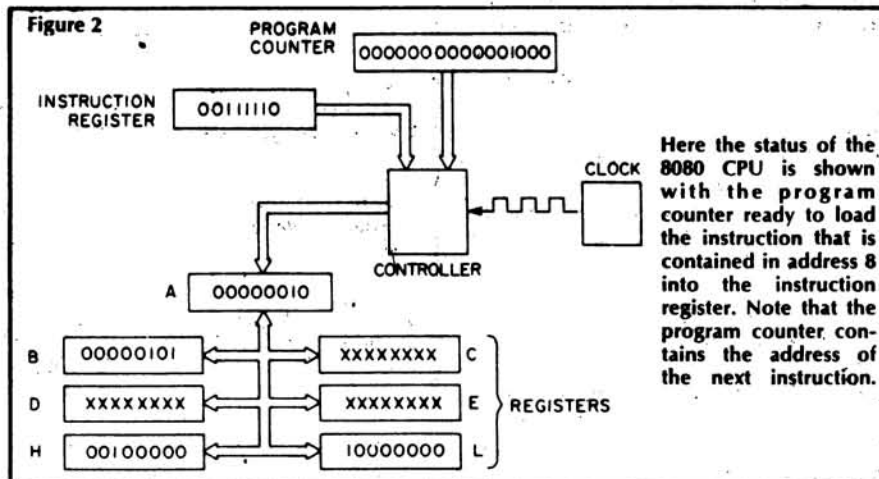
Answers. The question now is what to do with the answer. In our case, we have chosen to store it in a location in memory for future reference by the computer. The MOV_M,A instruction, at position 9 in our program, causes the contents of the A register to be moved to that location in memory *whose address is stored in the H and L registers!*

The final instruction is the halt instruction (HLT) which stops program execution. If it were not there, the program counter would merrily continue to step through all available memory—sometimes with intriguing results.

This program could be located anywhere in available memory that did not interfere with other machine operations. Our address designations merely reflect the sequential order of the program. To load it, we must first load the program counter with the starting address, then press RUN. The machine then takes over, driven and timed by the system clock.

It should be noted that our example was an addition routine. This was because practically all the computer does is add, with binary subtraction being merely a modified form of addition.

For example, multiplication can be



thought of as the successive addition of one number a certain number of times. One illustrative but inefficient way to do this in machine code for 5×8 would be to clear the B register and put 5 into the A register. We then add 8 to the B register 5 times each time we add 8 to the B register, we subtract 1 from register A. There is an instruction that looks for a zero condition in the A register. When this is the case, we have added 8 five times, or multiplied 5×8 .

For programmers wishing to use machine instructions, it is not necessary to memorize all the binary number associated with the machine instructions. There are programs called *assemblers* which allow one to write a program using the written mnemonics and numbers. The assembler then reads this written program called a *source listing*, and translates the letters and numbers into the binary information for direct use by the computer. An assembler code source listing of our addition program would look like that in figure 3.

When this is read by the assembler, the comments are ignored and everything else is translated into the binary code as shown above at a starting ad-

Figure 3

Opcode	Operand	Comment
LXI H	200 040	Put address in H & L
MVI B	005	Put 5 in register B
MVI A	002	Put 2 in register A
ADD B		Add B to A
MOV _M A		Put A into memory location
HLT		Halt

This is an assembler source code listing for our addition program. The assembler reads the letters and numbers and then translates them into binary numbers.

dress specified by the programmer.

The kind of machine programming we have discussed in this article is important to programmers wishing to write especially efficient or special-purpose programs. For the rest of us, it is perhaps interesting to realize that every time a BASIC program says $A=7+2$, something like that described above takes place in the machine . . . at 2 million cycles per second.

Computer Graphics

□ The CRT terminal is well known to computer enthusiasts as a fast and economical means of communicating with their machine. Mostly, the terminal accepts and displays information in the form of letters and numbers which are translated into ASCII (American Standard Code for Information Interchange) that the machine understands.

But the CRT screen, which is also used in television, is capable of displaying more than just alphanumeric information. Many home computers are supplied with a built-in graphics capability and many more have accessories available with which graphics can be added by the owner.

TV/CRT. How does the CRT termi-

nal display information, and how can it be made to do all the spectacular graphics that are becoming increasingly common in today's home computers? The cathode ray tube in a computer terminal is no different than the one in a TV set, so theoretically it should be able to display the same sort of picture information as a TV. In fact, many

Inside Computers

CRT displays are converted television receivers, and some personal computers output a video signal which can be sent via an RF modulator directly into the home TV set.

An electron beam is swept back and forth across the screen from top to bottom in just the same way as in a TV. Many displays meant especially for computers increase the sweep frequencies to give the display a greater "bandwidth," enabling them to display more information, but the principles are exactly the same. This means a CRT terminal requires vertical and horizontal sweep circuits just like a television.

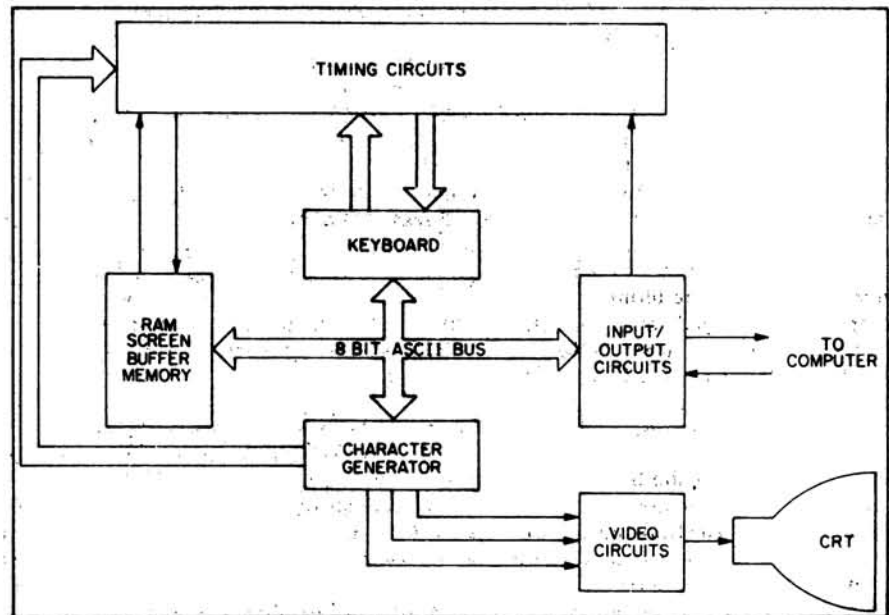
It also means that a video signal is required to deliver information to the display. The video signal is usually supplied by a circuit board inside the terminal, but in some cases it is generated by a video display board connected to the computer's bus.

In either case, the similarity to television ends here. When we get to the point of modulating, superimposing information onto that video signal, we enter a totally digital world. In television, the strength of the electron beam varies continuously as it sweeps across the screen giving black, white, and varying shades of grey. In color TV it's three electron beams for red, green, and blue. But in a computer display, that beam is either on or off; there are no shades of grey.

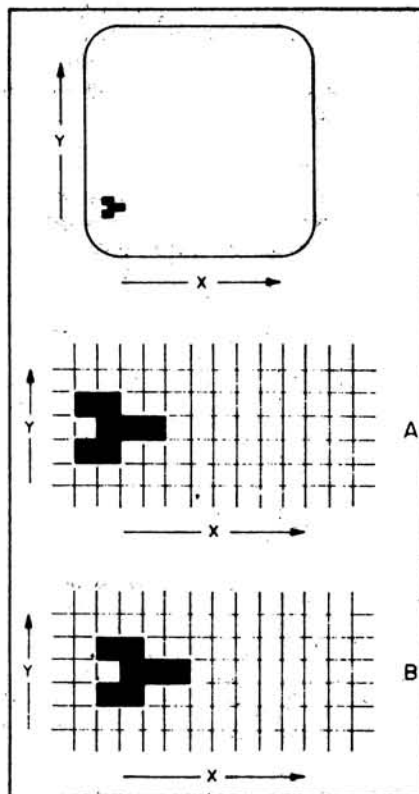
If we assume that our video circuits are exactly the same as a television set, we will have 525 scan lines per frame. What the computer does is supply information to turn the beam on and off at selected times so it will paint letters, numbers, or pictures on the screen. And that's really all there is to understanding computer graphics.

Screen Buffer. I'm kidding, of course, but in a sense it is true. The CRT displays information, and any information it does display must be in the computer's memory in binary form. The memory where alphanumeric and graphic information is stored is called the *screen buffer*, or sometimes the *graphics buffer*. A buffer is any distinct area of computer memory which is set aside for a specific purpose, in this case, the storage of display information. Most terminals today are at least partially "intelligent," which means they have a memory area for storing display information that is separate from the computer's main memory.

As it sweeps the electron beam across the screen, the terminal repeatedly looks at the information in the screen buffer



This block diagram shows the specialized circuitry used for graphics generation. Note the screen buffer; this stores the contents of the screen while allowing local updates.



A computer moves a rocket across a screen by turning on the neighboring set of cells.

to know what it is to display, when to turn the beam on and off. If the information in memory changes, the display on the screen will change the next time

the terminal looks at that location.

In order to understand how all the stuff the computer has to display is organized, we have to think of the screen as a matrix with X and Y axes. If you are thinking of buying a terminal, this will be an important concept to know. You will want to know how much information your terminal will be able to display at any one time.

For displaying text, 24 lines of 80 characters each is considered a good size. But terminals with 16 lines of 64 characters, 16 lines of 32 characters, and so on are also available with corresponding differences in price.

Resolution. When evaluating the graphics capability of a terminal, the matrix will refer to its *resolution*, that is, how many individual zones the screen is divided into which can be individually turned on and off. For home computers, a relatively high resolution screen may have something like 192 by 256, or 49,152 individual "cells" or "pixels," whereas a lower resolution display may have something like 48 by 100, or 4800 points. As we shall soon see, there doesn't have to be a location in memory for every point on the screen. One location can tell the terminal to turn on a certain area of the CRT display screen.

On a low-resolution display, the figures will look angular and blocky, while a high-resolution display will show finer detail and smoother curves. Of course, a high resolution display will require

more memory for its screen buffer and will cost more than low resolution, and for color graphics, even more memory space is required.

The method mentioned above of turning on a certain area of dots on the screen, such an area is referred to as a *dot matrix*. By using dot matrix, a computer terminal can generate a given set of letter and number characters. It can also generate a set of graphics characters and thereby save on the amount of memory it needs for fairly high resolution graphics. Here's how it works:

Characters. Inside the terminal, there is a special IC called the *character generator*. Actually, it is a ROM, or Read Only Memory, which contains codes for a given set of characters that the terminal is able to display. A dot matrix with a character in it is shown in Figure 2 along with the binary codes that turn on the proper dots.

Obviously, the electron beam can only write one row of a character on each sweep. Now remember, each cell location corresponds to an address in the screen buffer. If that cell is 8 lines high, the terminal looks at its address for 8 sweeps of the electron beam before going on to the next row.

Our good old character generator sends out 8 different bytes for each cell on the screen. They go to the same location in memory because the beam has to go by 8 times to display one character. So to paint a single character, our terminal has looked at one memory location 8 times, but each time its contents were updated.

O.K. that's all very detailed and it will do simply to think of the CRT as being capable of showing distinct areas

of dots that can be on or off in any combination. We can make letters and number with them as just described or we can create odd shapes.

Some personal computers, such as the Commodore PET and the Exidy Sorcerer, have an additional graphics mode where by pressing a shift key, the user can type a set of graphics characters, which are really just different dot combinations in the CRT's cells.

Saves Memory. There is always a tradeoff for this method. We get better graphics than if we only had the option of turning whole cells on or off. We can combine the graphics characters to form lines and curves at dot resolution, but we're limited to the combinations defined on our system. On the other hand, we've saved valuable memory which would have been required to give us control over every single dot.

But enough of electronics. Let's proceed to some concepts of how we use and program computer graphics. We'll assume for our purposes that our graphics system is one that does not use the special graphics characters described above, but is one with moderate, say 100 by 200, resolution.

There are a number of graphics languages and programs written in them on the market. There are also a number of versions of BASIC which have been modified to include commands for the control of the graphics. They all have two things in common: they must be able to turn a given cell on the screen on or off, and they must provide a means of referring to each cell by location.

Graphics in BASIC. Now comes the fun part. To demonstrate how a graph-

ics capability can work, we will use two hypothetical commands built into BASIC to draw a simple picture of a spaceship and then "fly" it across the screen. The commands are ON(X,Y) and OFF(X,Y) where X and Y stand for the screen coordinates of the cell location. ON and OFF simply means that that location is lit (ON) or not lit (OFF).

The first thing we do is turn on the cells we wish to represent our spaceship in its starting position. We have to name each cell in terms of its X and Y coordinates and tell the computer to turn it on. The computer then tells the CRT terminal to turn on the locations in its screen buffer that will light the corresponding areas on the screen.

We then see displayed on the left side of the screen our little spaceship. How do we make it fly? Simple. We tell the computer to turn off the left-most cell in each row of our picture (that is, of our spaceship) and turn on the next cell to the right. When it has done this the spaceship will appear to have moved one increment to the right.

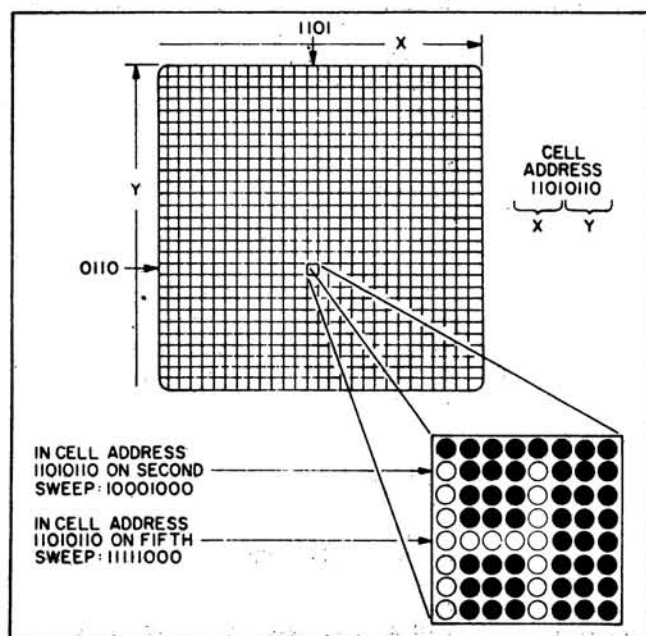
This is because when the terminal scans its screen buffer it will now see that the cells we want turned on have all shifted one place to the right. If we repeat this process by successively adding 1 to the X coordinates of those cells we want lit, our spaceship will appear to move across the screen.

We can control the speed of the spaceship by having the computer wait some fractions of a second before changing the contents of the screen buffer. Since the CRT displays 30 frames per second, the same as a television, motion can be very rapid and smooth, or very slow.

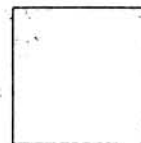
Lines and Curves. In the same way, we can draw a line or a curve on the screen by turning on cells or dots in succession and leaving them on. Some graphics systems allow you to specify a direction in degrees from a starting point and the number of "steps" you want your line to take. In this way, pictures and geometric figures can be formed conveniently.

One graphics system uses the concept of the *naki*. The naki is a little mouse or turtle which can move in specified directions on the screen. It can be told to move leaving a "trail" or line, or to move without drawing a line. For instance, the following short program would draw a square with its sides the length N.

```
10 INPUT N
20 DRAW
30 AHEAD
40 RIGHT 90
50 GO TO 30
60 END
```



A video terminal screen is divided into thousands of cells which are in turn divided into dots. A cell is called a character matrix. The one illustrated here is an '8-by-8' matrix. The matrix, which is identified by an address, is scanned eight times, once for each row of dots.



Inside Computers

Very complex figures can be drawn by modifying and combining these simple commands, and by utilizing the computer's ability to repeat operations many times. In the case above, the computer would go on drawing the same square until it was turned off.

Some home computers provide for control of their graphics by means of

a joystick. This consists of two potentiometers which represent the X and Y axes. The computer reads the different voltages output by different positions of the potentiometers. These are then referenced to the locations in the screen buffer memory which are loaded (turned on or off).

Computer graphics is an exciting way

to present information and it is appearing with increasing frequency as an integral part of home computer systems. One of the best comments on the value of graphics for conveying information is the name of a San Francisco Bay Area graphics consulting firm which calls itself "1K Graphics." . . . because a picture is worth a thousand words. ■

Computer Communications

□ MOST PEOPLE WHO HAVE begun investigating the world of computers are probably familiar with the most common means of communicating with the machine—the keyboard terminal. This type of input output, or I/O, device reads the different keystrokes typed by the operator and converts each into a distinct pattern of 1's and 0's which are then transmitted to the computer through one of its I/O ports. A port is simply a location (sometimes a memory address) where the CPU looks for incoming information from the terminal and to which it transfers data to be read by the terminal.

For the computer to receive meaningful data from a keyboard terminal, the operator must first formulate it in terms of the alphanumeric characters available on the keyboard and the computer must be programmed to interpret the symbols sent to it. This works quite well for a great many applications, but what if we want the computer to do something that requires so much constant input that the task of formulating it and typing it in all the time would be too dull and repetitive for a human to put up with all the time?

Computer Control. Let us take for an example, the task of reading—and later controlling—the temperature in some device. While that job could be done by a simple thermostat, we will see that using a computer will enable us to expand incredibly the possibilities for using the temperature data it reads.

The first problem is to get the information into the machine. We could have an operator read a thermometer and type the readings in via the keyboard, but he would soon get frustrated, especially if we wanted readings every half second or so. With the computer, we can do it automatically. This whole subject of automatic data gathering and external control is called *process control*. It is simply obtaining information

from the real world, processing that information according to a program in the computer's memory and performing operations based on that information.

This means that an incredible variety of specialized I/O devices and programs must be devised and tailored to the specific application. These can range from sensors of all kinds to stepper motors, valves and electric trains. We will now look at some of the general methods that are used to interface a computer to the outside world.

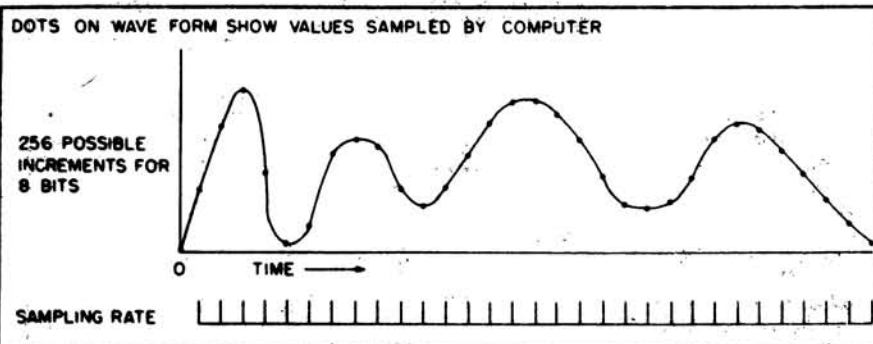
Analog to Digital. Since the real world does not normally present convenient digital data, some method must be found to translate *analog* signals into digital data. The general device for doing this is called an analog to digital, or A/D converter. An analog signal such as an audio wave, has constantly changing values so the best a digital device can hope to do is to *sample* positions of the wave and to approximate its shape in the form of digital information. Given the speeds at which computers operate, this can be done with surprising accuracy.

Let us look at the wave form shown here. It is applied as a continuously varying voltage to the input of the A/D

converter which is at one of the computer's input ports. The computer then reads the instantaneous value at regular intervals and accepts the digital value through its port. It can thus form an approximation of the wave form. The greater the frequency of this *sampling rate*, the more accurate the digital approximation will be.

Each sample that is read by the computer represents the value of the analog signal at a given point of time. For an eight-bit A/D converter, this value can be resolved into 256 increments ($2^8 = 256$). For example, it would be possible for such a device to measure a range of temperatures from 0 degrees to 100 degrees Celsius to an accuracy of better than 0.4 degrees Celsius! This accuracy can be enhanced by more complex (and more expensive) 10- or 16-bit converters.

The opposite of A/D conversion, digital to analog conversion, is also done, often on the same circuit board known, naturally, as an ADAC. Here, the digital information is output at a specific rate to a network of resistors, filters and op amps to produce an analog signal. The most sophisticated D/A converters are used in computer music



A computer can represent a wave form by sampling the level of the wave at a given point in time. The computer can interpolate the wave in between the points. The audio industry is just starting to use this same technique to make master recordings at the studio. The record albums that are produced are clearer and have more range than analog recordings.

applications where the highest quality complex waveforms are essential.

One inexpensive example of these principles, now available to hobbyists, is the Data-Bag™ speech processor made by Mimic Electronics (P.O. Box 921, Acton, MA 01720). This device allows the user to digitally record his voice in the computer's memory and then play it back by having the computer output those memory contents through the D/A section of the Data-Bag to a speaker. In all this there is a tradeoff the user must be aware of. The faster the sampling rate, and hence the more accurate the representation of the wave form, the more information must be stored—and the more memory must be taken up with stored data.

Of course, most real world applications are concerned with more than just inputting and outputting the same information. The real power of the computer, after all, is to process the information and make decisions based on it. And the true purpose of real-world connections is to enable the computer to act independently on those decisions. In this context, then, they are merely the things which implement the end results of the processing which is the computer's main task.

Control. By outputting analog signals, the machine could be used to control conventional analog devices. Let us take a look at our example of temperature control. The A/D converter, under software control, samples and presents temperature readings to the computer which reads and compares this data with its program information. If the temperature is outside the specified parameters, the computer outputs a digital value to the D/A converter which ap-

plies the proper analog voltage to the motorized temperature control until the readings again satisfy the limits.

Well, we certainly don't have to invest in a computer simply to use it as a thermostat! As I said and will emphasize again, the power of the computer lies in its speed and its ability to analyze large amounts of data and make decisions based upon it. So let us set our machine a somewhat more challenging task.

Multiple Data. We have a chemical process in which the computer must simultaneously monitor several thermometers, flow meters and pH meters, control various valves and temperature controls, and, in addition, prepare a human-report on materials used, temperature variations and the like. In addition, it must be alert for critical conditions (or combinations of conditions) and sound an alarm, and shut down operations if necessary. It should be apparent that the main problem here is not so much in the various on-off or analog devices at the machine's ports, but more in the considerations put into the software program.

Having the computer monitor many different inputs simultaneously is accomplished with an analog multiplexer. A typical such device can monitor 16 channels by rapidly scanning the inputs. In much the same way, an automobile distributor can service 8 different spark plugs seemingly at the same time.

Parts of the computer's program can then store the different values in memory locations allotted to the different input channels. This information can then be manipulated analysed and output in whatever manner the programmer thinks best.

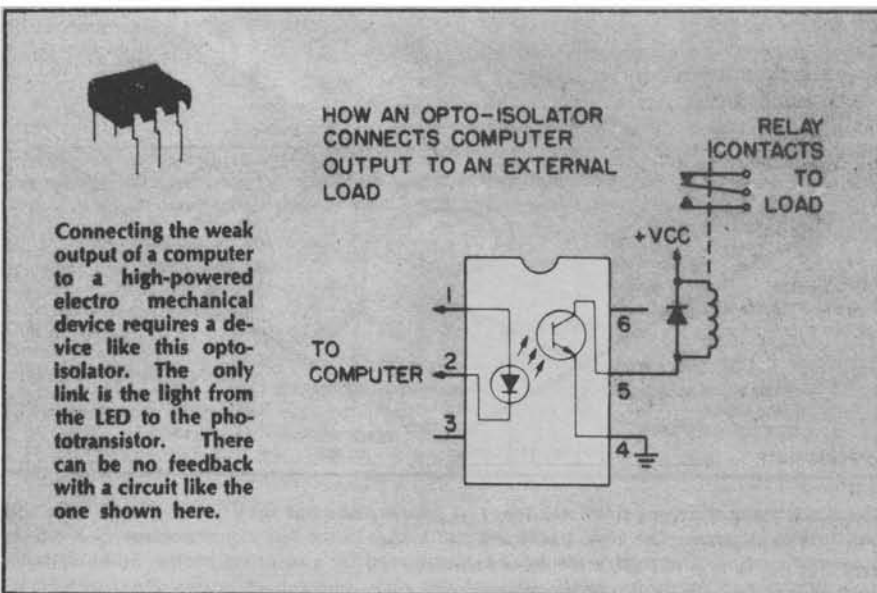
There are two more devices that are very important in process control applications. The first is the *real time clock*. Since all computer functions are controlled by various timing circuits in "clocked logic" there is a central timing reference in the form of a crystal oscillator known as the system clock. Its typical frequencies in home computers are between 2 and 4 MHZ. Since a crystal controlled clock is extremely accurate, divider circuits can be added to give the system an internal source of time reference calibrated in hours, minutes, seconds and fractions of seconds. The computer can then be programmed to refer to this time standard in order to time and delay the external processes it is controlling.

The other important device is one which enables the computer to turn on and off high voltage or high power devices safely. Obviously, the low voltage levels present at a computer's output port are not capable of controlling large devices by themselves. One solution is to use them to turn relays on and off which in turn controls the devices that do the actual work. The problem with relays and transistor switches is that there is danger of high voltage getting back into the computer and destroying components. The solution for simple on-off functions is to isolate the voltage with an *opto-isolator*. As the diagram shows, this is a device consisting of a photo transistor and an LED. The LED is connected to the computer output and is totally electrically isolated from the phototransistor. The phototransistor is turned on by the light emitted when the LED is on. It can thus be used to turn on any type of relay or other device while protecting the computer circuits from dangerous voltage.

This is the type of device we would use to sound alarms and activate sprinkler systems if our computer detected dangerous conditions, in our chemical process. Of course, with the computer in charge of things that should not happen.

If it did, the operators could use the computer to analyse the recorded data and the records of its control functions to generate a report on the entire process.

And that brings us back to the point I have been emphasizing. The program control and analysis of the data are the crucial aspects. The external devices must sense accurately and carry out the machine's orders exactly. But the whole system is still merely a tool which is used to extend the power of the human mind. ■



Operating Systems

□ Eventually, every enterprising home computerist, if he has been using cassette storage for any length of time, will look at the system he started out with and begin dreaming of upgrading it to do even more wondrous things. Visions of printers, X-Y plotters, color graphics, and controlling external devices will dance in the heads of those who have become relatively competent at programming and operating their basic systems.

Disk Drives. One of the first and most popular upgrades to a personal computer is, of course, the addition of a floppy disk drive. It is also one of the most dramatic. A disk system really makes a computer come alive and seem more "magical" after those interminable waits for a program to load from cassette. It also makes the system a serious contender for use as a business tool.

The big advantage of a disk system is the increased storage obtained and the speed with which stored programs and data can be read and utilized. We hear much about single and double density, access time, formats, etc., that is, the physical capacity of the disk itself. But one major consideration is often neglected. It takes a very sophisticated computer program, that is, software, to even begin to use the capabilities of a floppy disk system.

Operating Systems. Such a program, called an *operating system*, is far beyond the abilities of most home computer users to write and is therefore usually supplied by the manufacturer or by an independent software house. Anyone who is contemplating upgrading his system to include floppy disks should look well to the type of operating system that will be available to him. Also, by examining what an operating system is and what it does, we will get a better idea of some of the true potentials of today's microcomputers.

Essentially, the operating system is a program which supervises and coordinates the operation of all the various programs and peripheral devices (printers, disk drives, terminals, etc.) operating on the computer system. It keeps track of the order in which things are done, the names of different programs in memory, their locations, and it allows the user to tailor the system's operation to optimally suit his needs.

To understand how the operating system does this, we will need to use the concepts of "devices" and "files." A device can be any peripheral device connected to the computer: a CRT terminal, a printer, a disk drive, and X-Y plotter, etc. The operating system communicates with and uses these devices by means of special sub-programs known as *device drivers*.

Different peripherals require different software routines to drive them and an operating system may not have drivers for all possible devices. Nonetheless, a set of such routines is essential if the computer is to communicate with external contraptions, be they teletypewriters or robot arms. The ambitious programmer who wishes to run some exotic gizmo with his computer usually has to deal himself with writing the software to run it as well as building the hardware. A good operating system, though, will provide an easy method for the user to add device drivers of his own making.

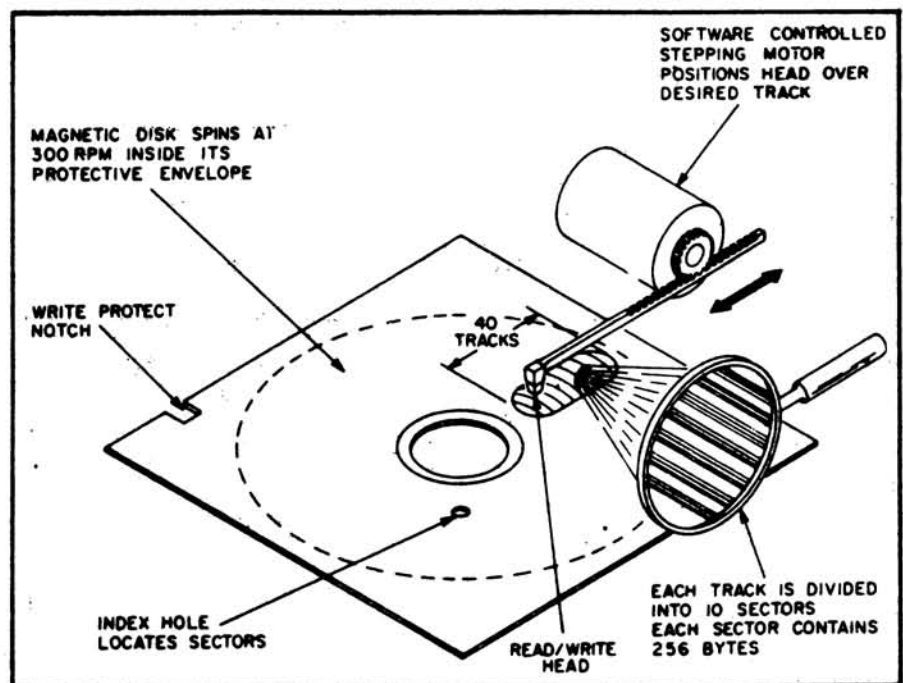
Files. The concept of a "file" is a little more complicated. To explain it, we will use the example of the floppy disk. The

floppy disk is both a "device" which requires its own device driver, and a gigantic "file" itself. It is the place where all the other blocks of data and programs, or "files" are kept.

In fact, you can think of the disk drive, the device, as the filing cabinet, the physical storage unit. All the pieces of paper and records in the filing cabinet, then, correspond to the data blocks or files, recorded on the disk. The way the data is stored on the disk is illustrated in the accompanying diagram.

The floppy disk is a circular piece of material much like magnetic recording tape. It rotates inside a protective envelope under the read/write head of the disk drive. The disk is divided into a number of concentric tracks, and each track into a number of sectors. A sector is the smallest unit of information the disk can access at one time. In a typical home computer diskette, each sector contains 256 bytes and each track contains ten sectors. On a diskette with 40 tracks, that adds up to over 102,000 bytes of storage!

With the disk rotating continuously



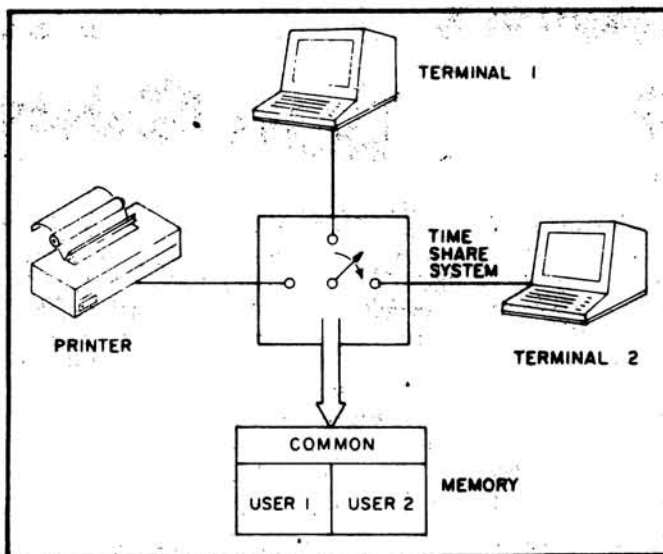
There are many different sizes and types of floppy disks but they all look much like the one in this diagram. The disk tracks are not visible lines but rather sectors of a single magnetic coating. The read/write head is positioned by a stepping motor. Some systems have read/write heads for both sides of the disk. Index hole marks the first sector.

under it, the read/write head moves quickly in and out until it has found the track requested by the computer. It then waits over that track until the proper sector comes around. At that point the head "taps" the disk and either reads data from it into the computer or writes data onto the disk.

File Finder. The job of the operating system in all of this is to keep track of where all the things are on the disk so they can be found when requested by a program. The operating system acts much the same as a poor, harassed file clerk who is constantly asked to fetch and file records. To ease its task, it generally uses part of the disk to establish a *directory* for itself. The directory is a little reference file the operating system uses to tell itself where everything is stored on the disk.

The name of each file (a file can consist of as many sectors as necessary, all grouped under one name) is chosen by the user and generally means something in English, such as STARTRK or BIOTHM. The operating system reads the file name when it is entered on the terminal, then looks it up on the disk directory. From the directory it finds out what track and sector it must go to to find the beginning of the file. We might also liken this process to finding a book in the library. First we go to the card catalog (the disk directory) and find out what aisle and shelf (track and sector) the book is on, then we go get the book.

So far, we have seen the actions of the operating system with the main means of mass storage, the disk drive. Let us now explore some of the things that are made possible by a typical microcomputer operating system. Not all the methods described here work for all operating systems because the approach of different authors varies. The



In a time sharing system various terminals, printers and other "slow" peripherals share the use of a single computer to enable more efficient use of the machine. This requires a special operating system to keep the different inputs and programs sorted.

general concepts, however, apply to all operating systems.

One major thing we can do with an operating system is take a file and transfer it from one device to another, or copy it from one disk to another. Many operating systems contain a program called PIP (Peripheral Interchange Program), which simply copies a file from one device to another. Using this program, a businessman can record the status of his inventory or receipts on a disk kept outside his shop and update that backup disk, say, every week. Thus, if anything should happen to his daily working disk, he only has to go back one week in his paper records to get things straightened out. A better idea is to update the backup disk daily—with PIP it only takes a few seconds.

An operating system increases in versatility if it is able to treat certain peripheral devices as if they were files. In that way it can use them as sources of data to be copied to a corresponding

file on the disk and later used by a program. For instance, if a program is monitoring a temperature at some peripheral, that information can be read as a file by the operating system and stored as a file on the disk. It can later be read as a file from the disk and copied as a file to a printer or used as a data file by another program.

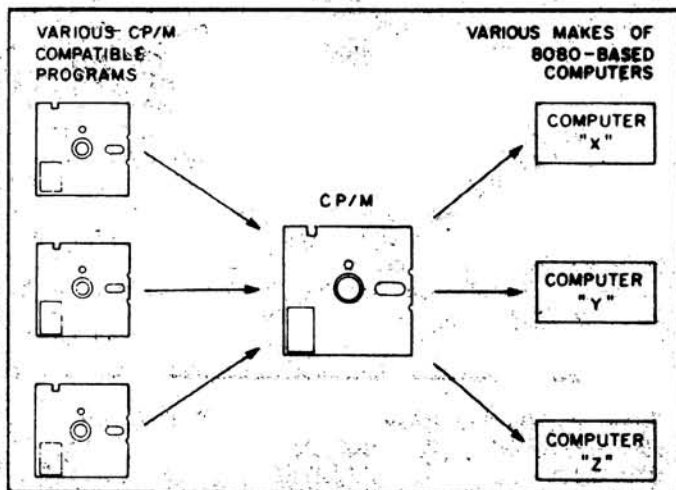
With various peripheral devices connected to the computer (CRT, printer, disk, etc.) the operating system must always be aware of which one is really running the show, that is, which terminal is allowed to give commands to the operating system, tell it to shut up, power down, or fetch a file. This terminal is known as the *console terminal*.

In most cases, the console terminal is the CRT terminal since many printers don't even have keyboards. It is, however possible to have more than one console terminal and more than one user program working on the same computer system at the same time. This is done by means of a method called *time sharing*, and it is one of the most powerful features available on today's computers.

Thumb Twiddling. Peripheral devices, especially those with mechanical parts, are terribly slow when compared to the speed of a microprocessor CPU. While the printer is busy printing the letter "G," the CPU is spending most of its time twiddling its electronic thumbs waiting until the printer is ready for it to send the next character to be printed.

A time sharing system makes use of these idle moments by paying attention to other console terminals. To do this, it "polls" them one after another in much the same way an automobile distributor services spark plugs in rapid

The CP/M operating system can be modified slightly to run on any 8080-based microcomputer. Applications programs can then be written to conform to CP/M's requirements rather than those of the computer. Thus, any computer running CP/M can use any of the thousands of independently produced programs written to run under that operating system.



Inside Computers

succession. The operating system also has to assign certain portions of memory to each user and keep track of who is where and when.

In addition, the system must keep track of priorities: if two terminals are ready for its attention at the same time, it must decide "who's on first." Thus, in a time sharing computer, our operating system not only plays the role of file clerk and scheduler, but, it seems, that of diplomat.

Time shared systems on microcomputers seem to be the coming thing for small to medium sized businesses. Studies have shown that what a small businessman needs is a system with four terminals or less which sells for under \$10,000. And manufacturers are currently falling all over each other trying to meet that need.

A typical such system might be found in a parts store where inventory control is critical. One terminal might be out at the counter or in the shipping area. Sales and shipments are entered here and the computer then automatically subtracts the parts from the inventory file and adds the sale amount.

Back in the warehouse, there could be a terminal where incoming parts would be entered and added to the inventory file. Meanwhile, in the office, inventory reporting and bookkeeping could be

done using both files with a third terminal along with a printer. And all this could be (and is being) done by a single computer with the necessary memory, disk space, and an efficient operating system.

One thing must be noted, though. There is a distinct difference between the operating system and the various programs and computer languages which do the actual programs and calculations that the user finds useful. The operating system actually just does the housekeeping. But it represents something more, something very important.

The operating system provides a software interface between a computer language like BASIC or a large application program like "General Ledger" and the piece of hardware that is the actual computer. So we often speak of a language or a program that will run under a certain operating system. That means, no matter what sort of funny piece of hardware it may be, if a certain operating system has been adapted to run on it, any piece of software written to run on that machine also.

D.O.S. A number of manufacturers have produced Disk Operating Systems to run on their machines. Processor Technology has PTDOS, Heathkit has HDOS, Cromemco has CDOS, etc. There is also a very popular operating

system which has been modified to fit a wide variety of computers. It is called CP/M and was produced by Digital Research (P.O. Box 579, Pacific Grove, CA 93950). The reason CP/M (for Control Program/Micro) is so popular is that it has been adapted to run on a number of different machines.

Programs written by many independent software companies need then only be written in such a way that they run under CP/M and they can then be used on any computer that has a CP/M and they can then be used on any computer that has a CP/M operating system. This is a big incentive for independent producers of software to get busy and write programs that are CP/M compatible because they know they will find a large market.

Having more software available in turn makes the computer more versatile for the user, and increases its value to him. Expanding a personal computer from a hobby to a serious tool for business or career still represents a sizeable investment. The add-ons are themselves expensive, but the user should look beyond that to the features of the operating system he will be using—both at what he can do with the system itself and how much more of the world of micro-computer software it will open up to him. ■

Bubble Memories

□ The two types of memory technology most familiar to hobbyists and other small computer users are semiconductor and magnetic. RAMs, ROMs, EPROMs, and EAROMs fall under the former category; tape and disk storage are included in the latter. Most computerists have at least a passing knowledge of each.

There is, however, another class of memory that is beginning to emerge. This type of storage has a data access time, a cost per bit and a storage capacity that lie somewhere between those of the two familiar forms of memory. The new technology is called "magnetic bubble memory," and it's catching on fast.

Rising Bubbles. Bubbles have already made a name for themselves by finding application in word processing systems and telephone switching networks. Many of you have literally heard the results of bubbles in action. The taped

messages in the phone system most likely are "filed" by a bubble memory. Also, a telephone's time-of-day message is often harbored in a bubble. ("Good morning! At the tone the time will be ten (pause) thirty-two (pause) and forty seconds. (pause) Beep.")

Although still in its infancy, magnetic bubble storage is expected to account for \$50 million in revenue by 1981. By 1984, bubble technology should garner \$230 million; by 1986, a whopping \$688 million. Today about one-third of all bubble memory units are used by the military and aerospace industries. By 1983, it is predicted, fully one-half of bubble memories will be used in telephone systems, mostly in switching networks.

Although the price tag on these devices is somewhat high today, manufacturing costs are declining and, in the next couple of years, the price of magnetic bubble storage will dramati-

cally decrease. You can, therefore, expect bubbles to burst on the small computer scene very soon.

Early Developments. Magnetic bubble technology began in the Bell laboratories in 1967. Andrew H. Bobeck observed that small cylindrical magnetic domains form in thin films of synthetic ferrites of garnets when an external magnetic field is applied perpendicularly to the film's plane. He also showed that a varying field can move these domains laterally through the film.

A rectangular solid sheet of garnet material magnetizes readily at right angles to its surface, but not parallel to the surface. In the absence of a magnetic bias field (H_b), the material remains in a state of magnetic equilibrium; that is, the positive charge in the substance equals the negative charge, so that the net magnetic moment of the material is zero (Fig. 1). Magnetic balance is maintained even if the magnetic

bias field strength is increased a bit. This state is designated H_{B1} , and magnetic domains appear in serpentine, or strip, form within the crystal while H_{B1} bias is applied.

A second tier is reached when the magnetic bias field intensity is increased further. Here, the magnetic regions contract into cylinders (circles, if viewed from above). The diameters of the cylinders range from 2 to 20 micrometers, as a function of the material used and the strength of the applied bias. It is this H_{B2} state that magnetic bubbles are created and maintained, as shown in Fig. 2.

If the field intensity is further increased and maintained (H_{B3}), the bubbles disappear.

Obviously, bubbles are most stable under H_{B2} conditions, and such a magnetic bias field is supplied by a permanent magnet sandwiching the garnet bubble crystal. This also accounts for the non-volatility of the bubble memory. (A volatile memory loses its data when power is removed. RAMs are volatile memories. Non-volatile memories such as tapes and disks, retain their data even upon loss of power.)

A bubble or group of bubbles in isolation has only academic value. To make them useful as storage elements, they must be manipulated under controlled conditions. Bubbles can be made to move by applying a rotating magnetic field (a magnetic gradient) to the bubble structure, as shown in Fig. 3. When a surrounding pair of orthogonal field coils is energized, the bubbles move simultaneously, in order, through predetermined paths in the crystal structure.

These patterns, which usually take the form of chevrons, are thin film deposits on the surface of the garnet crystal. Upon application of 2-phase alternating current, a rotating field is created in the plane of the permalloy chevrons. The bubbles move along the path defined by the chevrons as a function of magnetic polarity (Fig. 4). Thus, a type of serial shift register is created where logic 0 and logic 1 are defined as the absence or the presence of a bubble, respectively.

Microcomputers First. It is expected that bubble memory systems will be used in microcomputer applications before they are used in large computers. Immediate applications exist for high-density bubble memory systems in smaller computers, while large computers will probably have to wait through most of the 1980's.

The microcomputer applications for which bubble memories are presently well-suited include data terminals, word

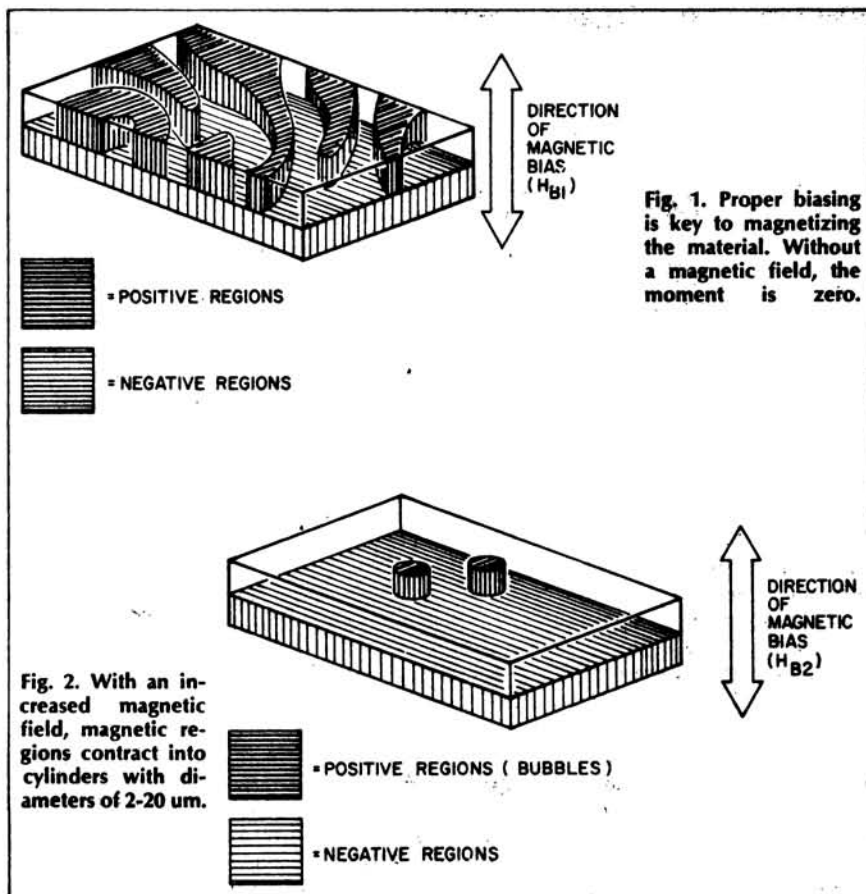


Fig. 1. Proper biasing is key to magnetizing the material. Without a magnetic field, the moment is zero.

Fig. 2. With an increased magnetic field, magnetic regions contract into cylinders with diameters of 2-20 μ m.

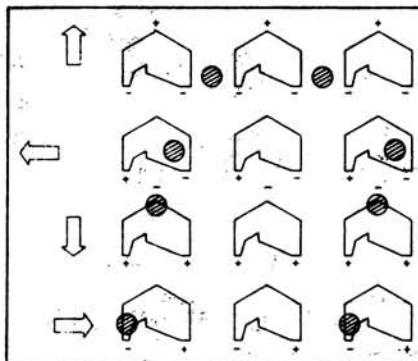


Fig. 3. By applying a rotating magnetic field, bubbles can be made to move. To be useful bubbles must be totally manipulated.

processing systems, industrial control systems, and telecommunication systems. These often require peripheral read/write memories storing from 128 Kbytes to around 2 Mbytes (the capacity range of most tape and diskette data storage peripherals).

With magnetic bubble memories, such storage can easily be built into the basic microcomputer system. Moreover, since they are solid-state, bubble memories can increase system reliability because of drastically reduced maintenance requirements.

Mass memory storage systems such

as tapes and disks are electromechanical, and are usually designed to operate in moderate environments. In contrast, bubble memories exhibit the high reliability associated with semiconductor systems. This reliability extends to the storage medium, too, because unlike tapes and disks, the magnetic bubble medium is sealed against contamination from the outside.

Data Storage Configurations. The magnetic bubble memory was originally configured to store data in a long-loop serial shift register manner. It turns out, however, that this scheme suffers from a few significant disadvantages. On the one hand, in order for any particular bit to be read, it must first circulate throughout the entire data loop. This takes time, a valuable commodity in data access. On the other hand, a single fault in the shift register architecture spells disaster for the entire bubble chip. The processing yield for the shift register technique is, therefore low, resulting in commensurately higher prices placed on the devices.

Major/Minor Loop. Another, more efficient, organization, called the major/minor loop structure, is a serial-parallel-serial technique in which data is serially generated in the major loop and moved over to the minor loops to

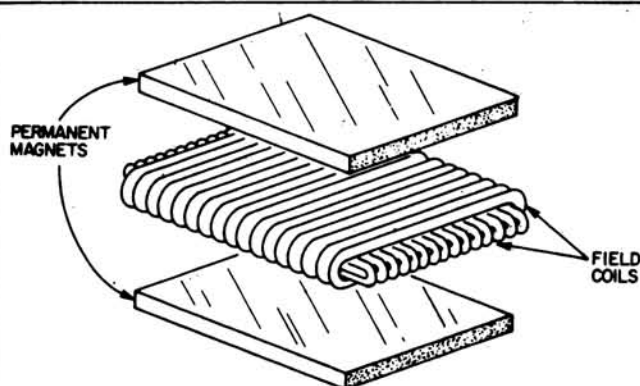


Fig. 4. Field coils and permanent determine the direction of the bubbles as they move. Thus, a serial shift register is created.

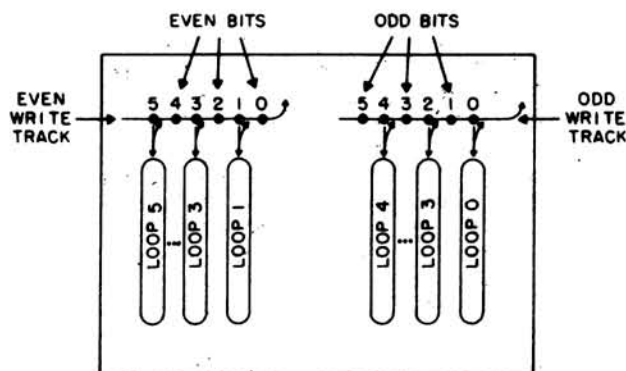


Fig. 6. The data storage area is divided into odd and even storage banks.

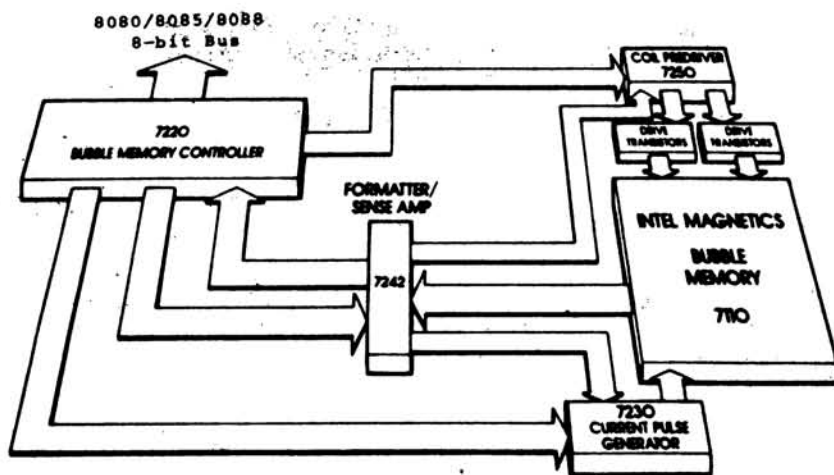


Fig. 7. Support large scale integration (LSI) circuits available with the Intel 7110 mega-bit bubble memory chip turn the memory module into a complete storage system.

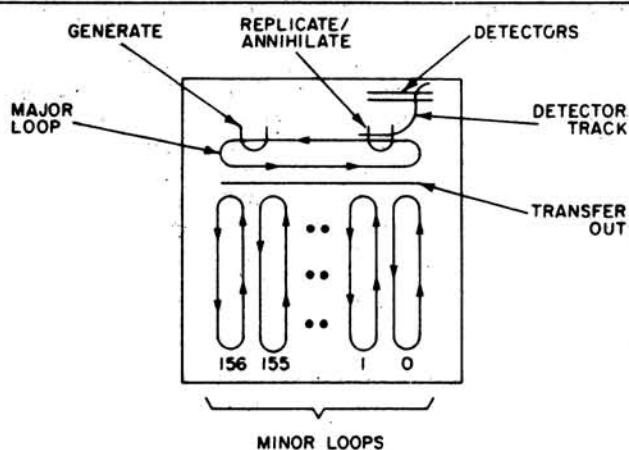


Fig. 5. This is a block diagram of the major/minor loop organization of Texas Instruments' TIB0203 92 kbit bubble memory. For reading, data is transferred from the minor to the major loops.

operation; the other involves the spacing between the minor loops, which also affects speed. Block-replicate organization remedies both problems.

In block-replicate organization, the major loop is split into an input track on one end of the minor loops, and an output track on the other end (Fig. 6). At the output end, each minor loop has its own replicate/transfer gate. During a read in this structure, the loops rotate to the bit position that corresponds to the desired data block. That block is then replicated onto the output track without the data ever leaving the minor loops. This replicated set of data then moves along the track to the detectors, and is read and destroyed—thus speeding up the read operations.

The second problem, that of the spacing between the minor loops, is also solved with the block-replicate method. The ends of these loops are separated by one bit position, so there is a space between each valid data bit that moves onto the major loop or is replicated in the output track. In a simple major/minor loop structure, these spaces must be read and ignored by the bubble memory controller. This cuts cutting the achievable data rate in half.

The cure for this problem involves splitting the data-storage area into odd and even banks. Each bank has its own generator and input track, but the generators are connected in series so that they create identical data on both tracks. Data moves along each track into position over the minor loops, but the odd bank has an extra bubble position, so when the first bit is opposite

finally be transferred, *en bloc* and in parallel, to the minor loops for ultimate storage. This is shown in Fig. 5.

The major/minor loop approach results in faster access times and, very importantly, it allows the inclusion of redundant minor loops. Redundancy permits defective loops (determined

during quality control procedures) to be eliminated from the memory map of the device. The bottom line is a fast, efficient and reliable chip.

As efficient as the major/minor loop structure is, however, it has two important drawbacks. First, there is a problem associated with a too slow read

loop 0 in the even bank, the second bit is opposite loop 1 in the odd bank. Then, when the data are shifted into the minor loops, every other bubble generated enters a loop.

The remaining bubbles—odd numbered bits from the even bank and even numbered bits from the odd bank—are shifted off the end of the track and destroyed. Because data is now generated in every bit position, the data rate equals the shift rate—twice as fast as in the simple major/minor loop structure. A further variation on the odd/even approach can make the data rate twice the shift rate.

Hardware Support. One of the major enigmas that faced manufacturers of bubble memory chips was that of developing hardware support for a working magnetic bubble memory system. Intel Magnetics (Santa Clara, CA) was recently successful in producing and marketing a megabit bubble memory.

Intel's 7110 bubble memory system is typical of the great strides that high technology has taken during the past decade. The 7110 bubble memory system establishes the new storage medium as a viable element in the computer memory hierarchy.

The 7110 bubble memory device it-

self is organized with block-replicate architecture. The device stores 2048 pages of 512 bits each, formed by combining two 256-bit registers with serial output channels. Each page is processed as 64 eight-bit bytes.

In a system operation, a page or burst of pages—up to the entire 2048—can be read or written at one system request. Pages are located so that after reading or writing one page, the next page is immediately available. The device can also be started and stopped between sequential pages to give controlled rapid access.

Based on a shift rate of 50 MHz, the average random access time of the 7110 is 40 milliseconds (ms); the maximum data rate is 100 kHz. A page-read or -write requires 327 shift cycles, so that the read/write time is 6.5 ms, and the average data rate is 78 kHz.

A block diagram of a basic 7110 system is shown in Fig. 7. Outside interface is provided by the 7220 bubble memory controller. The 7220 makes the system look like a peripheral to the microprocessor system bus.

The controller chip is a 40-pin device built with HMOS technology. It provides bus interface, generates all memory system timing and control, main-

tains memory address information and interprets and executes user requests for data transfers.

The 7242 formatter/sense amplifier is a dual-channel unit that interfaces with both channels of the bubble memory. It is a 20-pin device built with NMOS technology. It senses the low-level bubble signals, handles redundant loops, and buffers data. It also contains the burst error detection and correction circuits for each channel.

7110 Current Supply. The 22-pin 7230 current pulse generator, a Schottky bipolar TTL device, supplies the relatively high peak currents required by the bubble memory. It also contains a power-down circuit to shut off the current sources whenever the device is deselected. It has power-failure detect circuitry to shut down the bubble memory in an orderly manner.

High currents with peaks beyond the capacity of standard ICs are required to drive the coils. Therefore, the 7250 coil predriver interfaces with the bubble memory controller to driver transistors, such as quad bipolar transistor packs, or quad VMOS FET transistor packs. This CMOS device comes in a 16-pin package.

Floppy Disk Systems

□ Sooner or later just about every serious computer hobbyist reaches their computer's memory limit. There are some applications that require too much I/O action to be practical with a cassette based system or require too much memory space to fit in a RAM. It's this need for high-volume storage and rapid access time that makes a hobbyist disk system desirable. These systems don't come close to commercial hard disk systems in terms of performance but they can easily fill the computer hobbyist's memory bill.

Floppies have become the missing link to a midrange of random access memory systems. The floppy offers higher performance at lower cost than cassette and similar types of Input/Output (I/O) devices.

Well Packaged. The present standard floppy is an 8-inch flexible disk of a plastic material coated with a magnetic oxide. Looking a little like a popular 45 RPM record, it is sealed inside a jacket and there are no grooves on the surface.

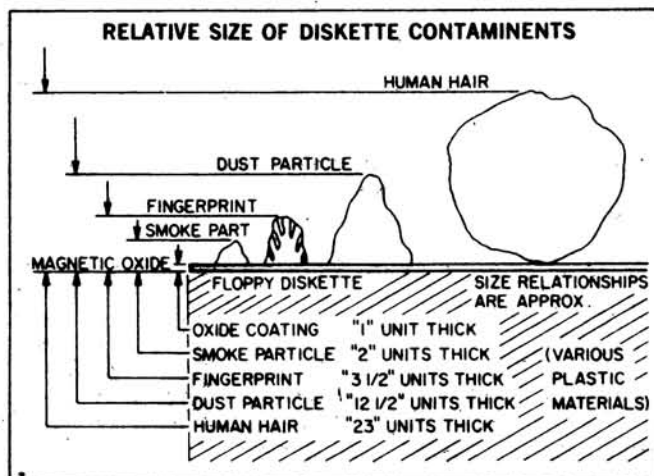
The disk cannot be removed from

the jacket which is designed to protect the recording surface. The disk is visible at a slot, a spot, and a hole in the center of the jacket. Users are told by the instructions that we must not touch exposed areas of the disk or write on it with anything firmer than a Q tip. Finally the user is admonished to return the jacketed disk to its outer envelope

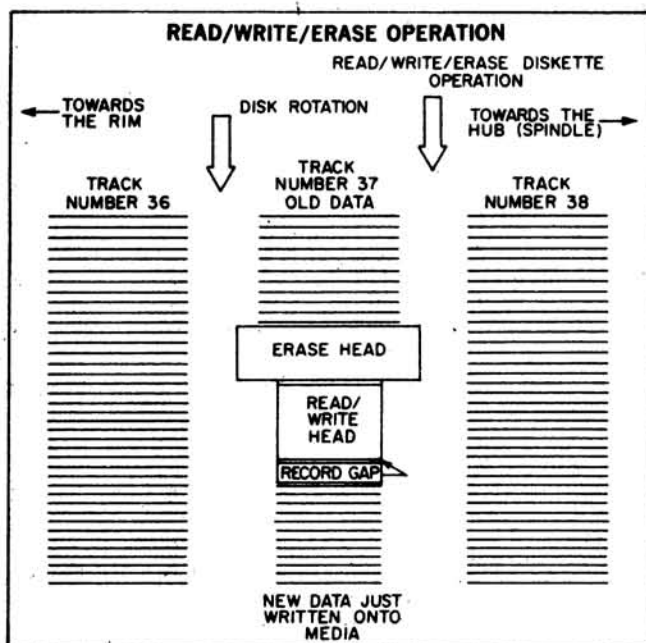
after they have finished using it.

The natural skin oil of fingerprints can damage the quality of music in needle and groove recordings, and in the super-miniature world of floppies, a fingerprint can destroy an entire segment of data. A dust particle can waste a dozen sectors and a human hair can reduce the effectiveness of the Read/

The metal oxide surface of a floppy disk is very sensitive to contamination. Each one of these contaminants (right) could cause a dropped bit. Since the disk is always kept in its protective envelope, this is usually not too much of a problem in daily usage.



Inside Computers



There is no physical track on the disk as with an LP record. The tracks are formed by the passage of the record (write) head over the surface of the disk. The disk itself appears to be a solid sheet of recording tape material.

Write/Erase (R/W/E) heads. The chart illustrates these comparisons.

It is understandable why the media, as the flexible disk is sometimes called, is permanently sealed with all of those implicit instructions. There is, however, an internal jacket wiper that continually cleans the rotating disk and removes contaminants, and floppies are reasonably rugged.

Hardware. The disk-drive hardware is add-on equipment to the main frame of the computer. Inside the drive are motors, driving mechanisms, and interfacing electronics that enable the drive unit to "talk" with the controller.

The diskette is inserted into the drive unit through a small door in the front. Once the door is shut, it is locked by the drive unit logic until the door release button is pushed to disable the drive assembly. The drive spindle centers and grasps the center of the diskette firmly as the motor comes up to speed. During power-up the diskette reaches a speed of 360 RPM, and R/W/E heads are stepped out to track 00 and a mechanical index hole provides the first location pulse for disk timing. In the IBM format this is the only reference to a physical location on the disk.

The floppy is firmly held against the recording surface, and the heads are positioned by a precise stepping motor. While the heads are positioned over the desired track, they ride above the spinning diskette. Once the correct track is located, in what is called a "seek operation," a head loading coil pulls the heads down onto the magnetic sur-

face. This operation is called loading the heads and it is controlled by a computer program.

When the heads are to be moved again, in another seek operation, to a new track, they are first unloaded, stepped to a new track, and then reloaded to a new track. This requires about three milliseconds per track of movement. It will further require about 50 milliseconds, per track, to move the heads and about 15 milliseconds of settling time.

Part of System. Floppies should always be thought of as a part of a

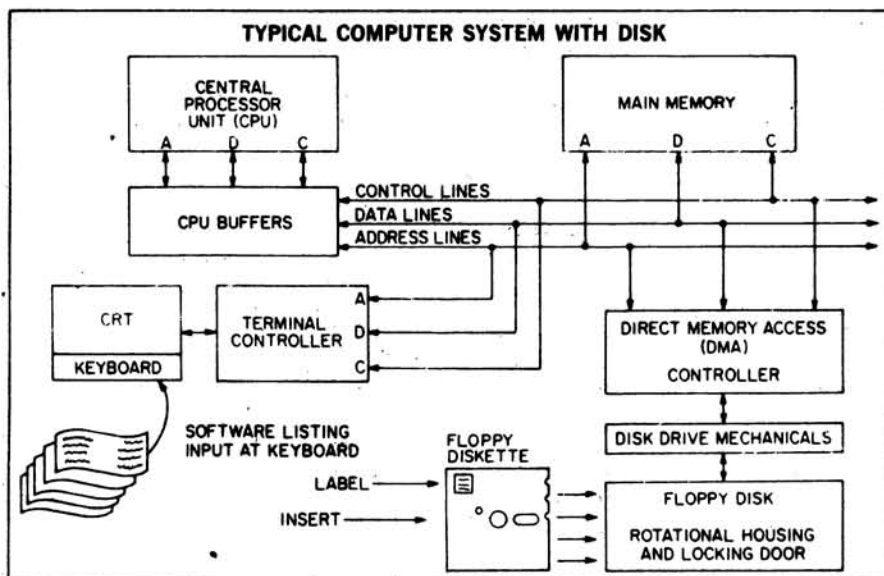
sophisticated cluster of mechanical drives, computer electronics, and software called a computer. Fundamentally this computer consists of a central processor unit (CPU), some memory, some interfacing devices, called controllers here, and an entry device such as a keyboard CRT, and some software.

Computer operation is made possible by a written program entered into the computer's memory and operated on by the internal microprocessor. The programs are fed by any one of a number of techniques: a paper tape punch, keyboard, cassette tape, or teletype. All of these are slow and time consuming. New ways are continually being introduced to feed the voracious appetite of computer memories. Floppies are the most versatile of the program instruction loading techniques.

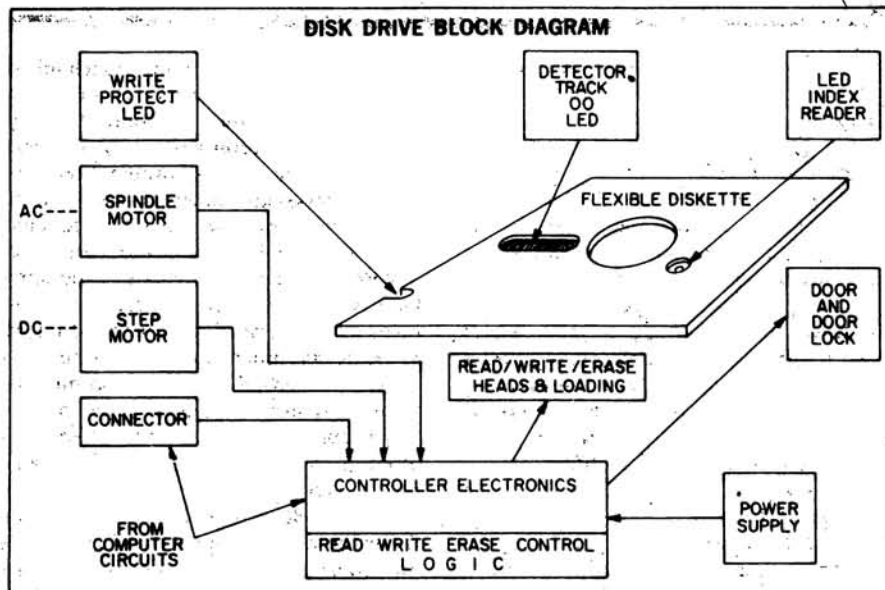
In a computer's time frame things go on a million times faster than in our brain's time frame. In such a whirlwind existence, telling computers what to do was a difficult problem. The answer was in the development of software which the computer could store in memory for reference each time it needed a new program instruction.

Software. The computer itself—the IC's, printed circuit boards, filter capacitors, chassis, and power supply—are all part of the hardware. All of this state-of-the-art electronics is just so much junk without a program, a way to make the computer compute. The instructions, written in computer languages, are called software.

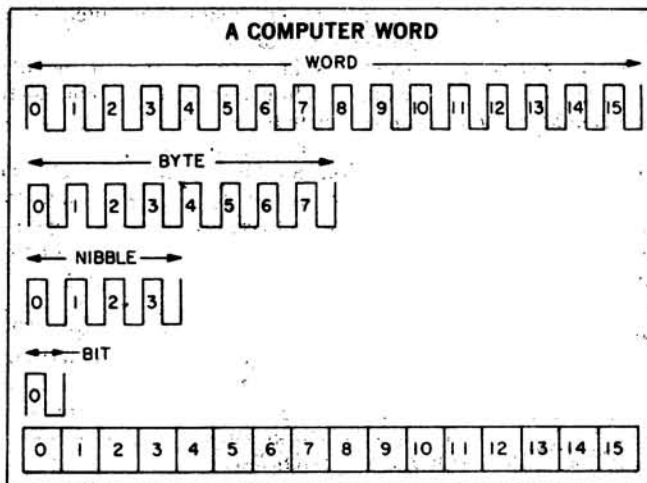
We talk with the floppy (human to



This block diagram of a typical computer system utilizing CRT readout and disk drive shows the parallel connections of the memory, drive controller and terminal controller.



You can easily see from this diagram how the physical configuration of the disk is related to the various electronic systems needed to control information storage and retrieval.



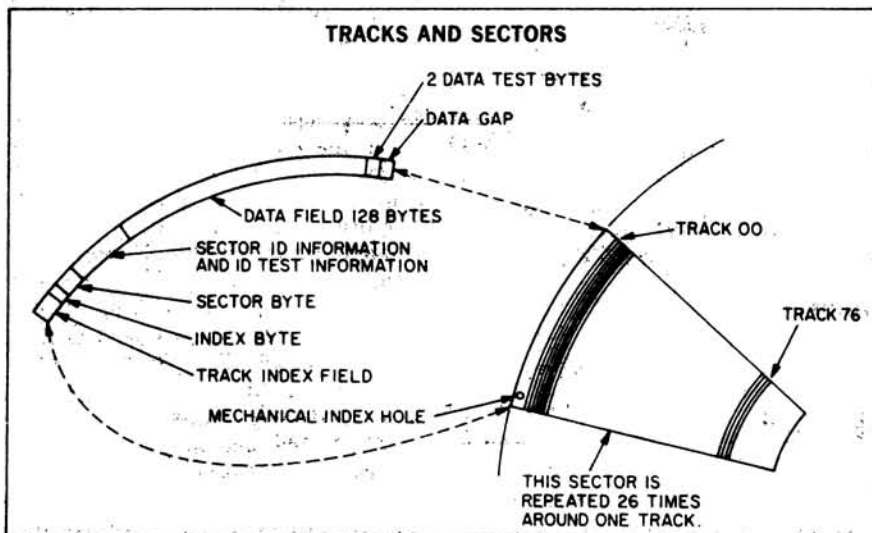
Software is the artful manipulation of the WORDS, bytes, nibbles, and bits utilized in transferring something meaningful between humans and machines without resulting in the total disorientation of the two parties involved.

floppy) through a series of interpreters comprised of software and hardware. First we place our instructions, using perhaps, Fortran in a machine assembly language. Located within this language and acting as a general interpreter, is a section of software called a file manager. Its job is to take general instructions, count words to be put on the floppy, calculate the number of sectors, tracks, and floppies that will be needed to store your file. All we do is to tell the file manager how much and where; it will do the rest. It will even put the data on the floppy then check to see if it got there, and if not correct its own errors. The special language of the computer is based on the numbers zero and one.

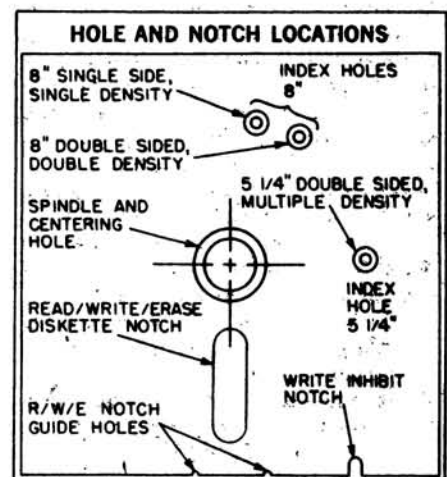
These two values comprise one "bit" of information. At this level there are no grey areas, no informational maybes. Facts are either a zero or they are a one. In this language a word has only one length for the micro-computer. It is 16-bits long. That is 16-bits having two states or thirty two pieces of information. However most peripheral devices, such as the floppy controller, that hunk of electronics that interfaces (talks) between the floppy and the CPU, is designed to speak in half words. This half word is called a *Byte*, pronounced "bite." A byte has 8-bits so there are two bytes to a full 16-bit word computer.

The story doesn't stop there. A new term is emerging in the industry as they learn to manipulate the byte. It is the half byte and is called a nibble.

Disk Mechanics. The R/W/E slot in the floppy jacket is two inches long by 1/2-inch wide. That little hole on the opposite side (8 inch diskette) is the mechanical index hole, which is a physical starting place (read by a LED



This diagram shows the relationship between track information and sector information as they are physically located upon the disk when utilizing the IBM-type disk formatting.



The disk's guide holes serve the same purpose as the arm on a record changer which positions the tonearm according to LP size.

Inside Computers

sensor) for recording information on the oxide surface.

On a tape cassette you can break off a plastic tab thereby preventing further over recording on that cassette. An industry term for this is 'record protect.' Floppy disk jackets have this same feature, in a slightly different form. On some jackets there is a write protect slot cut in the paper of the jacket. When not needed the slot is covered over. This slot permits a photo-optical system to shut off the write electronics when the system detects a slot. In this way valuable programs already recorded are not destroyed by writing over the program. With some drive units there is a switch to override this system, with other drives we must tape over the write protect slot.

The erase head most widely used may be either a tunnel-erase or a straddle-erase head. The tunnel-erase head design minimizes the influence of noise from data in adjacent tracks. This more clearly defines the erase band and improves the signal to noise ratio. Present usage seems to favor the tunnel-erase head design.

To place information on the diskette (to Write) and to retrieve information (to Read) a software plan called a *format* is employed to pre-organize diskette data fields. Where only a single reference is made to the mechanical index the resulting formatting is called soft sectoring. Most of the information presented here is for a single density, soft-sectored, IBM formatted diskette. After receiving the pulse that represents the index hole, the rest of the floppy is formatted from the software, or computer program.

IBM Format. The IBM 3470 formatted diskette, one of the more popular industry standards, has 77 tracks, with 26 sectors, (data spaces) formatted per track. There are 128 bytes per sector, 256 nibbles, or 1024 bits. The total numbers of sectors per single density diskette is 2002. Of the tracks, 74 are for data storage, two are set aside as alternate "bad tracks reserves," and an-

other track is reserved for maintenance purposes. A typical data transfer rate from floppy to controller is 250 kilobytes per second.

The tracks are numbered from the outside in with number 00 on the rim. Number 76 is the last track and is nearest the hub. Remember that there are no real tracks that you can see. They are the products of a software format, in this case IBM format 3470. These single sided, FM coded (more about that in a moment) disks can have a recording density of 3408 bits per inch.

A second type of sectoring is called hard-sectoring. Thirty two holes are cut in the diskette. These become the index marks for each sector. There is a 23% increase in data packing in hard-sectoring but the industry seems to prefer the soft-sectored format.

The sectors each contain a data field, with data gaps to guard this information, sector and track identification, again with gaps to protect this information, and guard bytes to further isolate sectors of information. All activity is in byte-length half-word groups.

Frequency Mod. The technique used to place data on the diskette is called Frequency Modulation (FM). Clock Pulses, 4 micro-seconds apart, from a 250-KHz clock generator, are placed on the floppy sectors forming data cells. This is the time from one clock pulse to the next. A magnetic transition within a data cell is read from the data

stream as a one; no transition as a zero. The data bit will fall into the center of the 4 microsecond data cell.

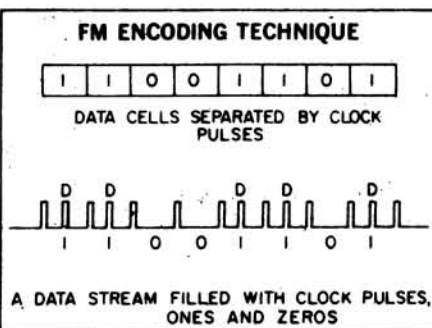
When the controller performs a read operation, the data stream going back to the drive electronics and on to the controller where the data is separated from the clock pulses, a known frequency, and also the index mark, sector data, and addressing information is removed before sending the data on to the Central processing unit.

As you might imagine there are sometimes errors in either a read or a write operation. A software test program is designed to search out these errors. They will be classed as either hard or soft errors. That is an error traceable to a piece of hardware such as a faulty motor is a hard error. An error traceable to a poorly formatted sector will be a soft error. One possible soft error would be a bad sector such that data could not be written into the sector. The controller would store the address of that sector in a *Bad-Sector* file in memory and search out another sector. Later if the computer addressed that bad-sector, the file manager would discover that it was bad and immediately go to the address of the new sector used in place of the damaged one. At some future date the user might want to replace the floppy if there are too many bad sectors.

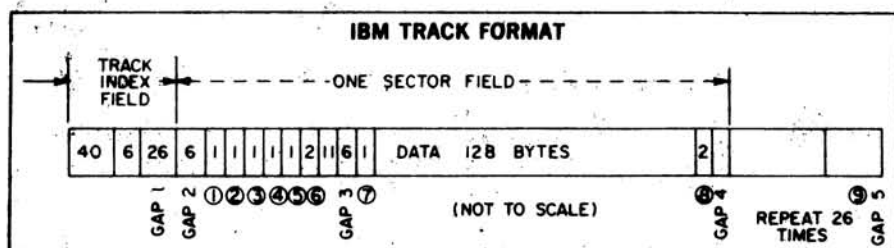
As the industry gets better at the techniques to pack and crunch data onto a small recording surface, floppy usage will increase. Shugart and Per-teck, among others, are offering the double density floppy employing a recording code permitting double data packing using the MFM or Modified Frequency Modulation code. Two other codes now in use are the Modified MFM or the Group Code Recording (GCR) technique. Double sided recording heads have also been introduced allowing recording on both sides of the floppy with transfer rates of 500 Kilobytes per second and 256 bytes per disk sector.

Mini-Floppy. Hardly had this double density, double sided floppy been introduced when the mini-floppy popped on the stage. Only 5¼ inches in diameter, it uses a coding scheme called the Modified MFM, with the index hole shifted 90 degrees to the right. The Radio Shack add-on TRS Mini-Disk System, uses a mini floppy.

As with all disk systems, a certain number of sectors are devoted to house-keeping, sometimes called labeling. This



The binary data pulse stream (lower line) is translated into and stored in the data cells (top line). 1's and 0's are shown.



IBM definitions: byte 1—address mark; 2—track #; 3—side #; 4—sector #; 5—sector size; 6—ID test bytes; 7—data mark; 8—data test bytes; 9—trailing gap.

usually consists of a directory, test programs, and other sector and track information so that it will talk smooth-

ly with its computer system.

The flexible diskette is a very useful peripheral device used with a I/O con-

troller and can be used to store special diagnostics, (hardware test programs) and debug routines. ■

Microcomputer Connection

□ One of the best software sources for your personal computer is someone else's computer data library. Whether it's another computer at the other end of town, or a commercial computer on the other side of the world, if you can get it to talk directly to your computer, you can swap programs and data at the touch of a button.

Personal computers talk to each other and to commercial time-share computer services through the telephone dial-up circuits; the very same wires that connect to your home or office telephone. Through a special device called a *modem*, the signals from your computer are converted to audio tones that represent the ASCII alphanumeric characters. These tones are coupled directly into the telephone wires, where they travel to the other computer just as signals from your voice would travel down the wires. At the receiving end, a similar though not identical modem converts the audio tones back to the electrical signals required by the computer.

Computer Talk. Because the transmitted and received computer signals are always in ASCII, it's also possible for the operators to communicate with each other by simply using the computers as rather expensive terminals. What is entered on the keyboard of one computer is shown on both computer displays, in much the same manner as a TTY or Telex connection.

Two practical examples will give you a better illustration of what's possible through a telephone connection. First, assume that your friend or co-worker across town—or across the country—has a program of particular interest to you. Rather than mailing or transporting a copy of the program on tape or disk, the two of you simply establish a communications link on the telephone circuit. This is, of course, done through modems (one must be an *answer* modem, the other an *originate* modem, or combination answer/originate modems can be used). At your request, your friend simply dumps the program into the modem, which in turn transmits it down the line to you. You see the program come up right on your display. At the end of the transmission you simply store the information from

your computer's buffer to the storage medium—tape or disk.

Our second example deals with a computer service such as MicroNet. You dial up MicroNet, and when you hear the computer's carrier tone in the telephone you connect the modem to the telephone line. You go through an identification procedure and then, using your computer as a terminal, tell MicroNet what you would like: news, bulletins, stored programs, or perhaps your own program that you have stored because it was too large for your personal computer's storage medium. Whatever, the computer service will download the data you have requested.

You can either process the data directly by using your computer in the terminal mode, or you can dump it to the buffer and then from there into your storage system. Your computer doesn't care how the data got into the buffer, so it processes the data that came down the telephone line the same as if you had entered it directly from the keyboard or from another storage medium (like a disk).

As far as using your computer as a terminal is concerned, there is plenty of software around that does just that. Often it's supplied with a modem that's specifically made for your computer. Other software permits the storage and transmission of computer signals. This is the easy part because it's handled by software. If there are any problems, they will usually be in getting your signal from the computer to the telephone line, and vice versa.

Using Modems. The purpose of this article is to get you using your computer for communications, so we'll concentrate on that and not the theory of how a modem works. Suffice to say that a modem is a device that converts a computer's numerical signal to audio tones and vice versa. Also, there are two basic modem types: *answer* and *originate*, which can be combined in a single device called an answer/originate or originate/answer modem. It's unimportant which does what. All you must remember is that an answer modem must always talk to an originate modem. If it's two personal computers it doesn't make any difference which is

answer and which is originate. (Some originate modems have a transmit/receive switch that allows it to converse with another originate modem, but the T/R switch must be thrown each time the computer receives signals.)

If you're "talking" to a computer service, you must use an originate modem, because the host computer service computer is always equipped with an answer modem.

Just for your information, the terms *answer* and *originate* come from large computer installations, where the computer was located in one location and the terminals were elsewhere, connected to the main computer through a telephone circuit. In the morning (usually) the terminal operator dialed the computer's phone number and the computer "answered" with a high pitched tone. The operator would then connect his terminal's modem to the telephone circuit and "originated" the data entry for the computer. Hence, we use *originate* modems at terminals and *answer* modems at the computer. Personal computer to personal computer communication is an exception to this convention and, as stated earlier, either can have the answer modem.

There are basically two different answer and originate modem types; and these are available in three configurations. A modem can be either acoustically coupled or directly coupled to the telephone line.

An acoustic modem has a cradle for the telephone handset. One end of the cradle has a microphone, the other end a miniature speaker (reproducer). After the operator hears the computer's answer tone in the telephone, he places the telephone handset in the cradle, where the tone signals into and out of the phone are coupled through the microphone and speaker (acoustic coupling). When the operator is finished, simply hanging up the phone breaks the telephone connection.

A direct-coupled modem does exactly what's implied; it couples the signals into and out of the computer through a direct wire connection. After the operator hears the computer's answer tone in the handset, he switches the modem *on line*. The direct wire con-

nection *seizes* the line and the operator can hang up the regular telephone. When the operator turns the modem off the line, seizure is released, just as it is when the telephone is hung up when using an acoustic modem.

Most modems intended for use with personal computers are RS-232, meaning they are wired for direct connection through a 25-pin "D" connector to a matching RS-232 I/O on the computer. (Some older modems were also intended for use with TTY *current loop* terminals and have both the RS-232 and 20 mA current loop connections.) You simply use a patch cord to connect the RS-232 connector on the computer to the matching connector on the modem. Whether the modem has an acoustic or direct connection to the telephone circuit, it has the same RS-232 connector.

Some of the commercial modems intended for use with time-share terminals provide a selector switch that allows either acoustic or direct connection; this is just an "extra" or "plus" feature of the modem and has no effect on its operation. When the modem is switched for direct connection the acoustic coupling elements are turned off, so extraneous ambient noises are not picked up by the modem.

TRS-80 Special Case. While virtually all modems are RS-232, there is a special exception for the Radio Shack TRS-80 Model I computer, which is available from aftermarket suppliers—meaning it's not made or sold by Radio Shack. To get an RS-232 I/O (input/output) for the Model I computer means buying an expansion interface and a serial interface.

The aftermarket modems for the Model I computer need neither the expansion nor serial interfaces: they con-

nect to the edge card connector on the rear of the CPU unit (the keyboard). These modems are directly connected to the phone line, with the software needed to operate the modem supplied with the device. Aftermarket modems for the Model I are available in originate-only, originate with a printer RS-232 "D" connector output, or originate/answer.

Once you've got your modem and the necessary software, if required, all you need do is connect the two *telcom* wires from the modem to the telephone circuit. Or, if you're using an acoustic coupler, extend the telephone to the computer's locations.

There are three types of commonly used "home" telephone wiring: terminal block, four-terminal plug-in jack, and modular plug (or modular jack). The plugs, jacks and adaptors you will need for connecting your computer to the phone lines are available from Radio Shack and telephone accessory stores.

If you have the most common type of plug-in phone, the one that uses a 4-terminal plug on the telephone's connecting wire, you simply need a three-outlet "mult adaptor" (or multiple adaptor) and a prewired extension cable. The "mult" is a cube with a four-prong plug and three four-prong jacks—it looks similar to an electrical "cube tap." You unplug the phone from the jack, install the "mult," plug the phone back into the "mult," and you're left with two extra four-prong jacks. You plug the extension into one of the two spare jacks on the computer.

If your modem came with a four-terminal output plug the extension must have a male connector on one end and a female on the other. You plug the male end of the extension into the "mult" and the modem into the free end. If you're using an acoustic modem,

simply plug a telephone into the free end of the extension. If you want to plug both a direct modem and extension phone into the extension cable you must first connect a "Y" adaptor.

As a general rule, if your modem makes a direct connection to the telephone line it makes operation a lot easier if you also install a telephone adjacent to the computer. It's not so much that you can dial directly from the computer location but you can hear when the computer on the other end

First-Class Modems. For those with unlimited funds to spend on a personal computer there are two types of "commercial" modems available for minicomputers. One is a "junior" version of a "data set." The telephone handset and its dial facility are routed through the modem. After the communications link has been established to the host computer, a switch on the modem makes a direct-to-line connection.

The other gold-plated special is an auto-dial modem presently available for only certain types of computers. It has several interesting features: for example, the operator enters the desired telephone number on the keyboard. The modem then dials the number, checks for answer tone, seizes the line, let's the operator know communications has been established with the host computer. If the computer is busy—as evidenced by a busy signal on the phone line—the modem can be programmed to redial until the link is established. The computer can also be programmed to provide directory dial-up—the operator enters a name or code, the computer selects the phone number from a directory, and the modem handles the communications link automatically.

Now that you know about modems and telephone connections, it's time to get your computer talking. ■

Keyboard Interfacing

□ A standard keyboard is probably the first peripheral that the microcomputer builder will add to a homebrew computer. In most cases where the microcomputerist buys a ready-built or kit-form computer, the computer company will tell you how to interface the keyboard—and you'd better follow their directions if you expect the software to work! But, if you are one of those brave souls who digs in

and tries to build a computer from S-100 boards (or those that you design yourself!), then you will need some information on how to interface the keyboard with your project.

Keyboard Codes. Microcomputer keyboards can be classified by the type of code that they produce. We will find some older machines, such as teletypewriters, that use obsolete codes (such as Baudot). But most micro-

computers will use the standard ASCII (American Standard Code for Information Interchange) keyboard.

The ASCII code uses seven bits to represent up to 128 characters or functions. In most microcomputers we will use the lower seven bits of a standard eight-bit input port to handle the ASCII data. The eighth bit, the highest order, will be used for a *strobe* or *data valid* pulse. The strobe

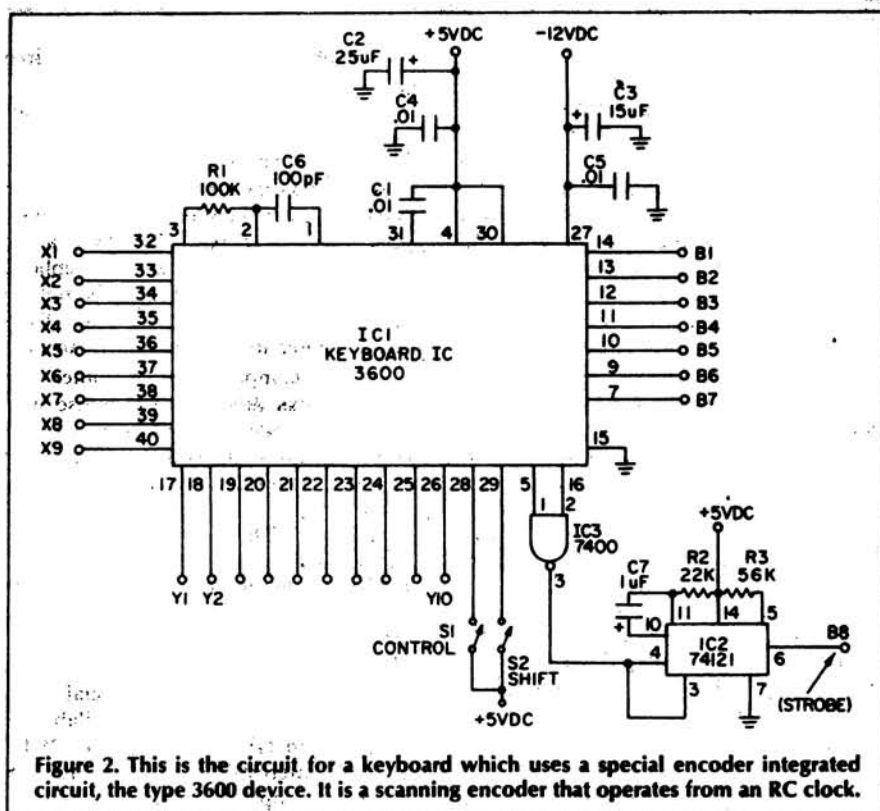


Figure 2. This is the circuit for a keyboard which uses a special encoder integrated circuit, the type 3600 device. It is a scanning encoder that operates from an RC clock.

pulse is used to tell the microcomputer that the data present on Bits B1-B7 forms a valid ASCII character. The code is shown in Fig. 1.

The circuit for a keyboard that uses a special encoder integrated circuit, the type 3600 device, is shown in Fig. 2. This device is a scanning keyboard encoder that operates from an RC clock. Resistor R1 and capacitor C6 form the timing elements for the keyboard. The seven output lines are to be connected to bits B1-B7 of the microcomputer input port.

The *strobe* bit is generated by a TTL monostable multivibrator (741-21). This circuit will produce one output pulse with a duration of approximately 1.5 microseconds. This speed is a little too slow for some microcomputers, but more about this a little later.

The selection of the ASCII character that will be output is determined by which button is pressed. The keyboard pushbuttons are connected to short lines of an X-Y matrix. On the 3600 encoder, there are nine "X" lines (X1-X9), and ten "Y" lines (Y1-Y10). This particular form of encoder requires pushbutton switches that actually connect the lines together, while others use capacitance keys or other technologies.

The *control* (CNTRL or CTRL) key and *shift* keys are SPST switches that will apply +5 volts to an appropriate input of the 3600.

There are two forms of strobe output seen on microcomputer keyboards: one-shot and constant. The one-shot consists of a short-duration pulse that "tickles" the microcomputer to let it know when the data is valid. This type of pulse is shown as pulse "B" in Fig. 3. The other form and you have to be aware of this type—is a constant level that remains as long as the key is depressed.

This type is shown as pulse "A" in Fig. 3. In many cases, it is almost irrelevant which is used! We can select a low-cost keyboard as long as it uses a constant level.

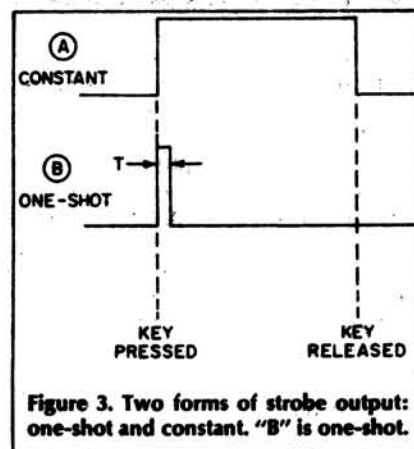


Figure 3. Two forms of strobe output: one-shot and constant. "B" is one-shot.

Figuring The Keyboard. Microcomputer hobbyists obtain keyboards from a variety of sources, and not all of them are too good about providing information on their devices. Generally, if you buy a keyboard from a reputable manufacturer, or a good vendor who normally serves microcomputer hobbyists, then adequate information will be forthcoming. Failing that, they usually have an in-house technical expert who can be called and asked questions.

But what happens to you when the keyboard is purchased second-hand, on the used/surplus market, or from a dealer who doesn't seem to care much about the customer? In that case, you could very easily become the "stuckee" in an expensive game! You will then have to figure out the keyboard yourself.

Unfortunately, there is no universally followed standard for keyboard features, how the features are implemented, or even pin-outs for the keyboard. The keyboard that you select might follow any one of several possible protocols!

Shooting In The Dark. First, try to obtain whatever printed technical data is available. If none came with the keyboard, then call the manufacturer.

Figure 1. These are the Hex codes used to transfer data from the keyboard to the main board. It is based on the ASCII system.

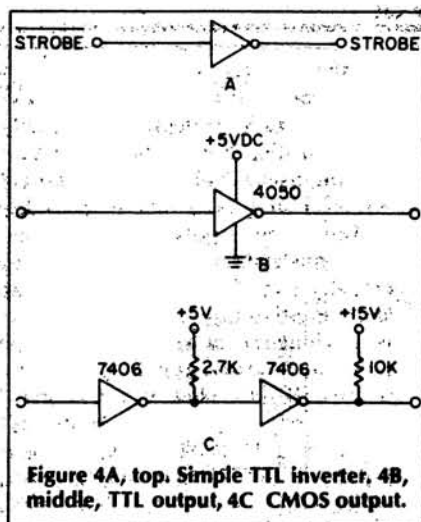
LOCATION	HEX CODE	COMMENTS
00 00	DB	IN 0
00 01	00	(data)
00 02	CB	TEST BIT 7,A
00 03	7F	
00 04	28	JUMP RELATIVE IF ZERO
00 05	FA	
00 06		
00 07		
00 08		
(DATA)		

I have found that calling the main number often results in your call being transferred to the sales department, regardless of what question you ask. I make it a habit to always ask to speak to the repair technicians, or the service department manager. These people usually have access to the schematics, and often do not mind sending copies.

You might also find the circuit for the keyboard in the service manual for the instrument that the keyboard came from. In my case, for example, I used a Texas Instruments Silent 700 keyboard for a couple of years. The TI service manual on that product contains the keyboard schematic, so I copied it for my own use.

The very first thing that you will need to know is the pin-outs of the keyboard connector. Most keyboards use a printed circuit "card-edge" connector or a 14-pin DIP socket. If the schematic is available, then use it to locate the various pins. Otherwise, you may be faced with the chore of manually dopping out the circuit. This job is not too fearsome if the keyboard uses the standard chip configuration; but it can be too tedious for words if it uses ordinary integrated circuits to generate the ASCII code and timing. Fig. 2 and an ohmmeter, or good eyesight, will tell you the pin-outs for those keyboards that use the standard chip.

The next thing to determine is whether or not the keyboard is TTL compatible. Some are not, especially on older machines like some teletype-writers. Most keyboards, however, are TTL compatible, so you will have

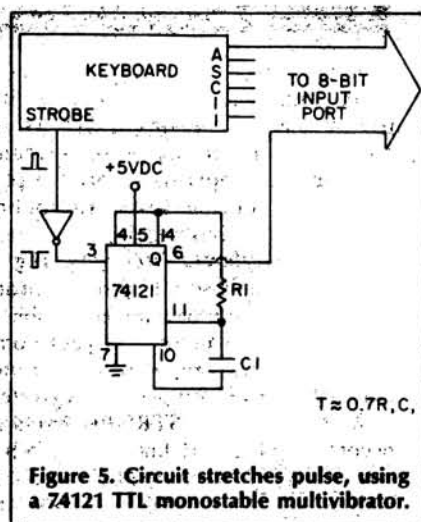


little to worry about. If the data bits come up with something between 2.5 and 5 volts when a key is pressed, it is most likely TTL compatible.

If it is not TTL compatible, or if you use one of those few microcomputers that needs a CMOS compatible keyboard, then some interfacing is needed. More on that in a moment.

The next thing that is needed is to know what the strobe looks like. Find out the duration of the strobe pulse (using an oscilloscope), and the polarity. It is also necessary to know whether the strobe pulse is one-shot or continuous. If the strobe is of the wrong polarity, which usually means that you got one that sits HIGH normally and then drops LOW, then use an ordinary TTL inverter to flip it.

In the case of my Silent 700 keyboard, the pulse was of the correct



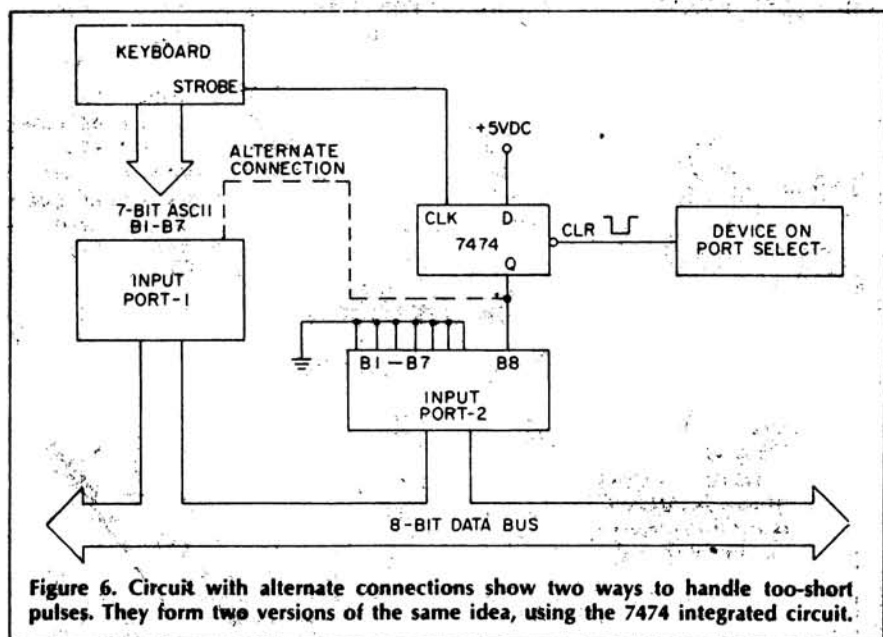
polarity (stays LOW and then goes HIGH when a key is pressed), but a friend of mine had the same model, from the same source, and found the strobe inverted. In that case, it was necessary to use an inverter IC to flip the pulse in the correct direction. Fortunately, there was an unused 7400 NAND gate section on one of the chips, and that could be wired as an inverter.

If the strobe pulse is too short, or if it is of the wrong type (i.e. one-shot vs continuous), then certain remedies are available and will be given shortly.

The last thing that you will have to determine is whether the data output of the keyboard is latched or unlatched. The latched type retains the last valid data that existed, i.e., the last character, until a new character is commanded. Other keyboards are unlatched, so they will either 1) have continuous invalid data ("trash") on the output lines all of the time, except during a key closure, or 2) return all output lines to ground during periods between key closures.

The type of keyboard that you have often determines whether or not the software works properly. On many computers, the program will loop continuously as long as the data on all lines is zero, and will try to input data whenever the input port sees anything except 00000000! If your keyboard is not the type that returns all output lines to ground between key closures, then the program will keep trying to input garbage data. Also, if your program uses a relatively slow loop (which might be fast to that computer, if the clock rate is slow!), then the duration of the strobe pulse may be too short!

In that case, the computer will seem



to miss some key closures! The cause of this annoying phenomenon is that the program is seeking a HIGH on the strobe line, but the 1 microsecond strobe pulse disappears before the program loop for the keyboard routine gets around to looking for it.

Solving Keyboard Problems. Fig. 4 shows the simple solutions to the problem of compatibility. If the strobe is a negative-going pulse, and your computer wants to see a positive going pulse, then use a simple TTL inverter, as in Fig. 4A. The STROBE line will become a STROBE line on the other side of the inverter (simple, huh?). This same is true if the other lines are "complemented," i.e. "turned over." You can use other inverters to flip that data as well.

But before you go solder-tacking a 14-pin DIP IC onto the circuit board in order to obtain a single inverter, make sure that there are no unused inverter sections of ICs on the keyboard already. You can also use sections of NAND (7400) and NOR (7402) gates for the same purpose—they can be inverter-connected very easily. Also, if you are planning to write your own monitor program, then it is a simple matter to have the program look for a LOW instead of a HIGH to indicate valid data.

The circuits shown in Figs. 4B and 4C are to make the keyboard TTL-compatible, or, CMOS compatible, respectively. If the keyboard seems to have CMOS outputs (10- to 15-volt data levels for HIGH), then use the circuit of Fig. 4B. This circuit uses the CMOS 4050 hex buffer IC, which will accept CMOS levels at the input. It will produce TTL-compatible output levels, provided that *only* +5 volts DC is supplied to the chip. If the chip power supply voltage is greater, then all bets are off!

In Fig. 4C we have the method for making the TTL keyboard compatible with the input of a CMOS system. The inverter is a 7406, and two sections are used in order to keep from flipping the data. The 7406 device has an open collector output, so requires a pull-up resistor to V+. The input section of the circuit uses a pull-up resistor to +5 volts (TTL supply), while the output 7406 section uses a pull-up resistor to the CMOS power supply (which could be as much as 15 to 18 volts DC). Again, use available sections of chips already on the printed circuit board, if any.

Pulse Stretcher. The problem of

too-short duration on the keyboard strobe pulse can be overcome using either of two strategies. The method shown in Fig. 5 is probably the simplest, and uses a 74121 TTL monostable multivibrator as a "pulse stretcher."

The name "pulse stretcher" is a little misleading, because it doesn't actually stretch anything: it is a one-shot multivibrator that generates a new (longer) pulse when the short pulse that needs stretching is applied to the input. The duration of the output pulse from a 74121 device is approximately $0.7R_1C_1$. Select values for R_1 and C_1 that will allow the strobe pulse to be HIGH long enough for the microcomputer that you are using to recognize it.

Most microcomputer keyboard routines are machine language programs that perform a loop looking for a HIGH condition on the strobe line. The programming guide for the microprocessor used in the computer will give the number of machine cycles, while the clock frequency gives the period of each. (Period = $1/\text{frequency}$).

If the clock is 1 MHz, for example, then the duration of each is $1/1,000,000$ or 1 μs . Similarly, if the clock is 2 MHz, then the duration of each pulse is 0.5 μs . We can then multiply the number of clock cycles needed to execute the loop by the time per cycle, and arrive at some idea of the minimum time required for the strobe to be recognized.

The seven ASCII lines from the keyboard are connected directly to bits B1-B7 of the computer input port,

while the strobe bit is first inverted (unless it is already negative-going) in a TTL inverter (7404, etc.). The output of the 7404 inverter is applied to the trigger input of the 74121, which wants to see a negative-going pulse. The output line from the 74121 is applied to the bit B8 line of the input port as the "stretched" strobe pulse. In general, a duration of several microseconds is usually sufficient.

Two alternate versions of the same idea are shown in Fig. 6, and form an alternate method for overcoming the problem of two-short strobe pulses. In this method, a Type D flip-flop is used as a 1-bit memory to remember the short strobe.

Let's review briefly the rules for the 7474 Type D flip-flop (FF). A Type D FF will pass the D input data (in this case a permanently wired HIGH) over to the Q output only when the clock (CLK) line is HIGH. In this case, the Q output will go HIGH and remain HIGH on receipt of the first clock pulse. But, there is also an active-LOW clear terminal on the 7474. All clear terminals cause the Q output to go LOW, while simultaneously causing the NOTQ to go HIGH. This terminal gives us a method for resetting the FF.

The seven bits from the ASCII output of the keyboard are connected to bits B1-B7 of an input port. The Q output of the 7474 FF is connected to either A) bit B8 of the same input port that handles the ASCII data (seemingly the most reasonable selection!), or, B) to bit B8 of another in-

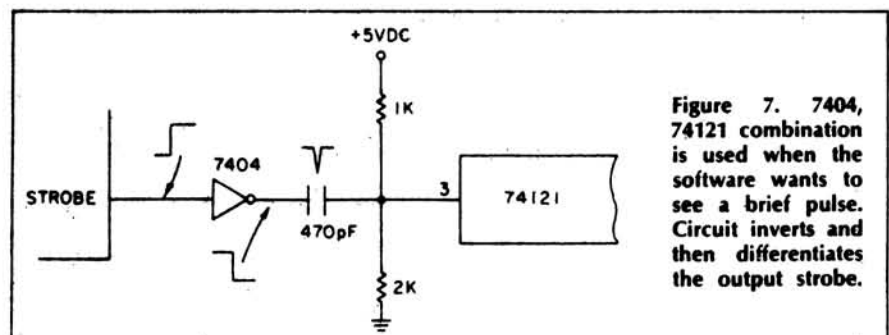


Figure 7. 7404, 74121 combination is used when the software wants to see a brief pulse. Circuit inverts and then differentiates the output strobe.

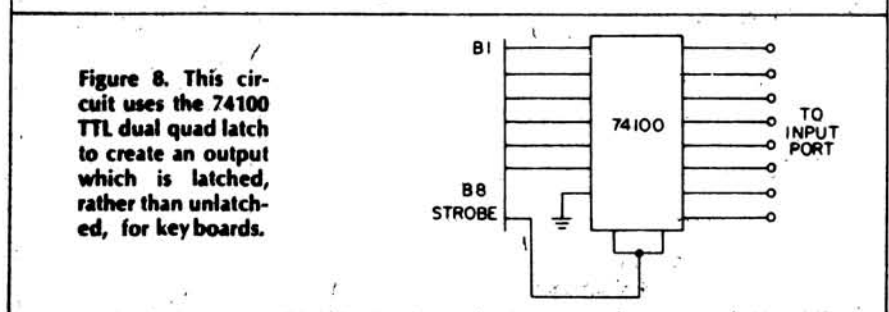


Figure 8. This circuit uses the 74100 TTL dual quad latch to create an output which is latched, rather than unlatched, for keyboards.

put port. The other bits of this second input port are connected permanently LOW. The program to input data will pole Input Port 2 for the binary number 10000000, and recognizes the number 00000000 as meaning that the data on input port no. 1 is invalid. But, when it sees 10000000, the data is known to be valid, so the program must branch to input the data on Input Port 1. This seems a waste, but is popular. If this method is used, then set bit B8 of the first input port permanently LOW by grounding.

After the program inputs the data from Port 1, then it must clear the 7474, or the program will never again see a valid data indication. There are several possible ways to do this. If there is an output port available, then write a loop into the program that will write a HIGH to one of the bits, and use that bit to control the 7474. This bit is normally kept HIGH, and is momentarily brought LOW to reset the flip-flop. The other method (shown in Fig. 6) is to connect the clear (CLR) line of the 7474 to the OUT line of an address decoder circuit.

We now know how to handle inverted strobes, and too-short strobes. But how do you handle the problem of the continuous strobe level (as opposed to pulse), when the software that you select wants to see a brief pulse? See Fig. 7.

The idea in Fig 7 is to invert and then differentiate the output strobe. The output of the keyboard strobe line will be a level that snaps HIGH when the data is valid. This level is inverted by a 7404, so that the input of the RC differentiator is a negative-going level

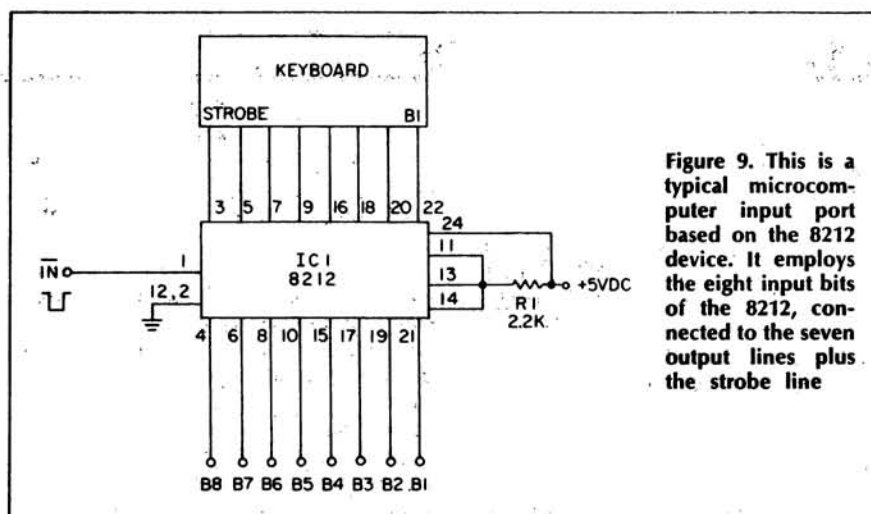


Figure 9. This is a typical microcomputer input port based on the 8212 device. It employs the eight input bits of the 8212, connected to the seven output lines plus the strobe line

translation. The purpose of a differentiator is to produce an output voltage that is proportional to the rate of change of the input signal, with a polarity that indicates the direction of the change. In this case, since the change is a very rapid, negative-going level translation, the output of the differentiator will be a high amplitude spike that is also negative. This spike momentarily causes the level of pin 3 of the 74121 to drop from its normal HIGH to a LOW condition, thereby triggering the monostable.

Using The 555. The 555 chip is often a good choice in this application because it is relatively easy to trigger, and can be operated at TTL levels. I recommend using an output buffer from the 555, however, because this device sometimes has a little trouble driving some TTL loads. I like to follow the 555 with either a 7404 or a

CMOS 4050 device. Note that the 7404 will invert the output, so two stages must be used.

The last problem that we see on some keyboards is the matter of unlatched output lines. We can create a latched output using the circuit of Fig. 8. The 74100 is a TTL dual quad-latch, so will handle 8-bits. The ASCII output lines of the keyboard are connected to the inputs of the 74100, while the 74100 outputs are connected to the lower seven-bits of the input port. The strobe line from the keyboard is connected to the strobe inputs of the 74100. When the data on the ASCII output lines of the keyboard is valid, then the strobe pulse will clock this data over to the outputs of the 74100. There, the data is held until another strobe pulse occurs, the next time that a character is selected.

OK, what what if you don't have a convenient input port? This would be the situation if all of the I/O ports are filled, you don't even have an I/O board, use a single-board computer, or, are homebrewing from a reu uP chip on a piece of perf-board. What do you do now? Make an input port, of course.

Fig. 9 shows a typical microcomputer input port based on the 8212 device. There are eight input bits on the 8212, and these are connected to the seven output lines plus the strobe line of the keyboard. The output lines of the 8212 are connected to the data bus lines of the microcomputer. It is very important that the correct IC be selected for this application; not all will work satisfactorily. The 8212 uses tri-state outputs, as must any input port IC. The output lines float at a high impedance that is neither a logical HIGH nor a logical LOW condition. This arrangement allows the 8212

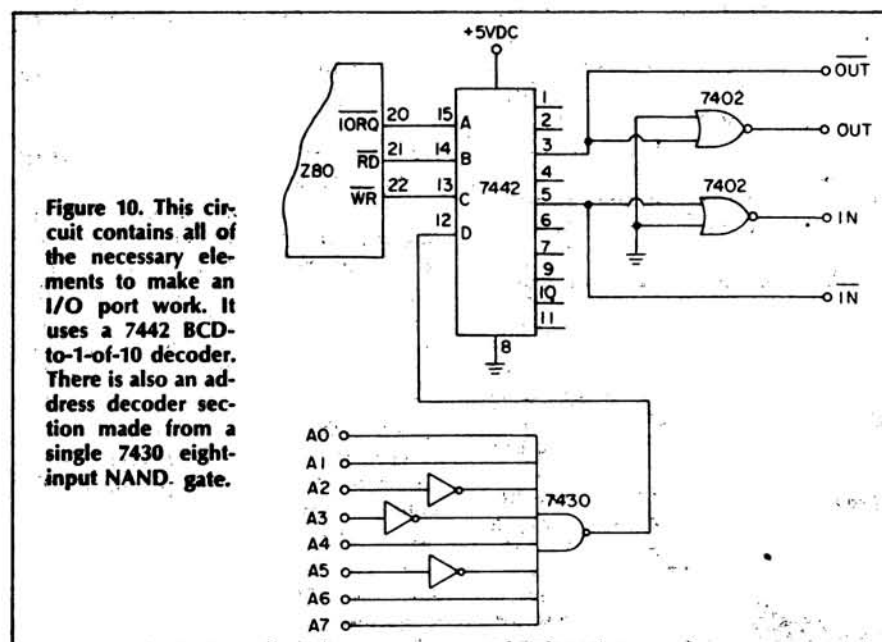


Figure 10. This circuit contains all of the necessary elements to make an I/O port work. It uses a 7442 BCD-to-1-of-10 decoder. There is also an address decoder section made from a single 7430 eight-input NAND gate.

Inside Computers

to float harmlessly across the lines of the data bus when it is not active. Otherwise, the input port IC would load the data bus lines and cause problems.

The 8212 has a terminal that will turn on the output lines when that terminal (pin 1) is brought LOW. This line is connected to the IN device select signal of the microcomputer. This wire will drop LOW only when the microcomputer executes a read (i.e. input) operation from that input port. Recall that we required a device select pulse in Fig. 6. The same sort of circuit will generate either type of device select pulse, and is shown in Fig. 10.

In the Z-80 device, there are three control signals that we need for I/O operations: input/output request (IORQ), read (RD), and write (WR). The IORQ is used in all input and output operations, while the WR is used only in output operations and the RD is for input operations. In summary, when the microprocessor chip is executing an input operation, both the IORQ and RD lines will drop LOW. Similarly, when the microprocessor is executing an output operation, the IORQ and WR lines drop LOW. It is these combinations that must be decoded for I/O operations.

There is also one further requirement. We cannot simply execute an I/O operation of either type: it must be directed *someplace*. That "place" is an I/O port with a specific number. In the Z-80, the address of the port called for in the program is passed along the lower order eight bits of the address bus (i.e. A0-A7). Since this address is eight-bits in length, it will access up to 2^8 , or 256, different I/O ports. We must, therefore, also provide an address decoder section in the device select circuit.

Figure 10 contains all of the necessary elements to make an I/O port

work. The heart of the device select circuit is a 7442 BCD-to-1-of-10 decoder. In this IC, there are ten output lines and four input lines that accept binary coded decimal (BCD) data. When the BCD word for one of the ten output lines is received, then that output line will drop LOW. For example, the BCD word 0010 has a decimal value of "2." When 0010 is applied to the ABCD inputs of the 7442, therefore, the "2-output" (pin 3) drops LOW.

The circuit in Fig. 10 uses an address decoder section that is made from a single 7430 eight-input TTL NAND gate. When all eight inputs of the 7430 are HIGH, then the output drops LOW. This means that the D-input of the 7442 will be LOW when the correct address is present on the lower eight bits of the address bus. We must conspire in this circuit to make all lines of the address bus HIGH when the correct address is present. This means that it will need inverters on those lines that are normally LOW when the correct address is present! In the case of Fig. 10, the correct address will be 11010011, so bits A2, A3 and A5 are LOW when the correct address is present. This means that we need 7404 inverters on these inputs of the 7430. Therefore, when address 11010011 is present, the 7430 sees 11111111.

When the program calls for a read (i.e., input) operation from port 211₁₀ (11010011₂), then the output of the address decoder is LOW, as are the IORQ and RD lines of the Z-80. The WR line is LOW. The way that we have this circuit designed, then, there will be a BCD word of 0100₂ (4₁₀) on the input of the 7442. The "4" output (pin no. 5) is connected to one input of the IN NOR gate (a 7402 unit). This line goes LOW, causing the 7402 output marked IN. This signal, then, is the IN signal that can be used at the input port. If an active-

LOW signal IN is needed instead, then use pin no. 5 of the 7442 directly.

A similar action occurs when the microprocessor chip executes a write (i.e. output) operation, except that WR is LOW and RD is HIGH. The BCD word that tells the 7442 to act in that case is 0010₂ (2₁₀). The "2" output is pin no. 3, so it will drop LOW when 0010 is present on the input lines. This line is used as the active-LOW OUT signal, while the output of another 7402 section is used as the OUT signal.

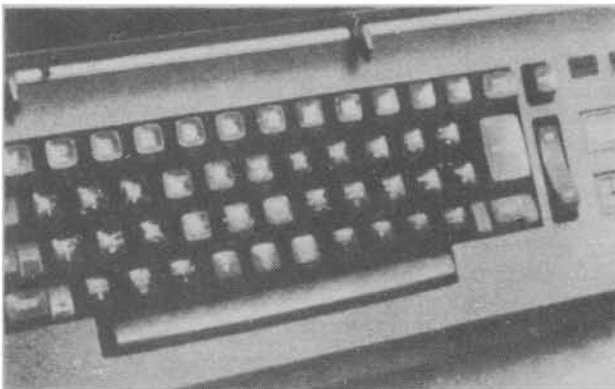
The 7402 can be replaced with inverters in many cases, but we can use the grounded inputs to control the ports. If we need a way to turn off any given port, then we can use these inputs, and bring them HIGH when it is necessary to turn off the OUT/IN lines. As long as this pair of inputs is LOW, then the OUT/IN lines remain in the LOW state.

Programming. It is probably impossible to write a program, or even a flow chart, that will satisfy all microcomputer users. We won't even try. In Fig. 11, however, is a Z-80 version that is used on a popular microcomputer. This computer connects the keyboard to input port-0, and the program executes a loop that inputs port-0, tests for bit B7 (the highest bit) for HIGH. If the bit is HIGH, then the program will fall through the next step until to the rest of the program. Otherwise, it will jump relative backwards to the first step, and inputs another byte from port-0.

This program is intended solely as an example, and not a universal. Keep in mind that the next step will be to transfer the input character to the portion of the program where it is required. This might be a video display memory location, or some subroutine.

We can also use the keyboard to interrupt the microprocessor. The interrupt method is often preferred because it allows the processor to do other work when there is no input from the keyboard. A typical microprocessor can do a lot of work in the 100 to 250 milliseconds it takes the fastest typist to key in data.

The strobe line of the keyboard is connected to the interrupt line of the computer. When the strobe becomes active, the interrupt service routine will be called and will transfer program control to the keyboard input program. This is probably the most efficient approach for most users.



Differences between characters appearing on a Selectric typewriter used as a line printer and those on a standard typewriter are shown here in this photo of a Micromatic-modified Selectric. The unusual keytop symbols are for computer codes.

Radio Receivers

What you will learn. When you have finished reading this article you will have learned what the parts of a radio receiver are, how they are similar to the parts of a radio transmitter, and how they separate the incoming radio frequency carrier waves (the carrier medium) from the audio frequency sound waves, which are the desired signal. You will also have learned that the important parts of the receiver are similar to the parts of a transmitter. Finally, you will have learned that CB Radio transceivers make use of this similarity of parts of the transmitter and receiver to make most CB sets into what are called transceivers.

A RADIO RECEIVER

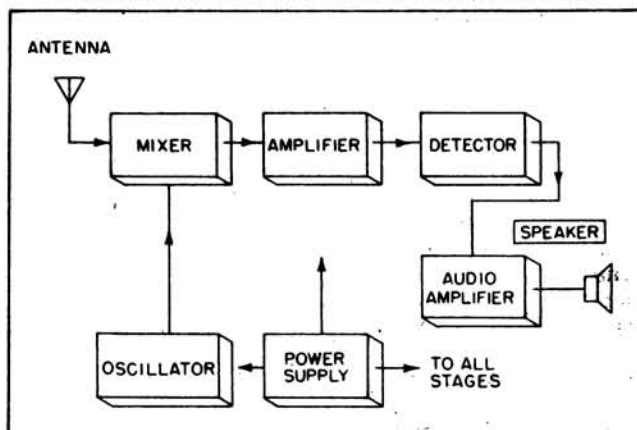
The block diagram for a radio receiver similar to the one in your home shows its major parts and how they feed the signal from antenna to loudspeaker.

The purpose of the radio receiver is to convert the amplitude modulation on the carrier back to its original sound. As the carrier increases in ever-widening circles on leaving the transmitter antenna—like ripples in a pool—its energy decreases in amplitude. The increasing circumference of the circles causes power in the waveform to be distributed over an ever-increasing area. By the time the signal reaches the receiver antenna it is rather weak, usually around a few thousandth or millionths of a volt. The receiver, therefore, must amplify the received signal to a level that will operate the speaker within the hearing range of the human ear. The receiver must also extract the audio component (the **envelope**) from the carrier. The carrier brings the signal to the receiver, but has no function in reproducing the audio frequency in the receiver.

The Power Supply—Each receiver has a power supply. Its purpose is to convert 115 volts AC from an electrical outlet to DC voltages that will operate the receiver properly.

The Antenna and Mixer—Radio frequency carrier waves from all stations within range of a receiver appear at the antenna of the receiver. When you turn the dial of your radio to a specific station, you adjust the electronic components of the **mixer** input so that the receiver will accept a particular carrier frequency and reject all others. The received carrier enters the mixer to be amplified. CB radio receivers, as well as other sets which need to be very sensitive because they must receive very weak signals (from weak transmitters, or from transmitters which are distant), must have an additional stage between the antenna and the mixer. This stage is called an RF (radio frequency) amplifier.

TYPICAL RADIO RECEIVER BLOCK DIAGRAM



QUESTIONS

- Q1. What part of the received radio wave does the receiver convert back into original sound?
- Q2. A radio wave decreases in power as the circumference of its area increases. What is the approximate amount of voltage that enters the receiver antenna?
- Q3. The _____ converts AC to DC voltages required to operate the receiver circuits.
- Q4. A single broadcast frequency appears at the input of the (antenna, mixer).
- Q5. CB radio receivers feed the incoming RF signal directly from the antenna to the mixer stage (true, false).

ANSWERS

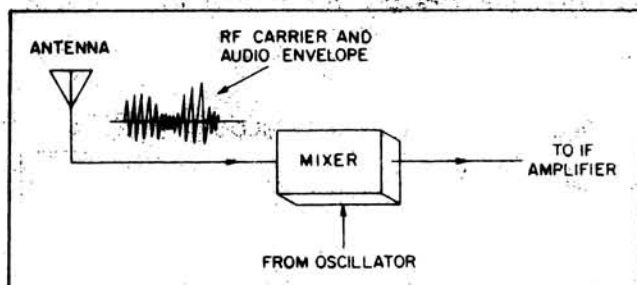
- A1. The **amplitude modulation** (or audio envelope).
- A2. **A few thousandths or millionths of a volt.**
- A3. The **power supply** converts AC to DC voltages required to operate the receiver circuits.
- A4. A single broadcast frequency appears at the input of the **mixer**.
- A5. CB receivers feed the incoming RF signal directly from the antenna to the mixer stage. **False.**

The Oscillator—The receiver oscillator is similar to its counterpart in the transmitter. Both generate a signal of constant frequency and amplitude. The purpose of the receiver oscillator is slightly different, however. It is designed to generate a frequency that is a constant number of kilocycles above the carrier frequency regardless of the transmitting frequency the receiver is tuned to. Tuning the dial changes the value of the electronic components in the frequency-generating part of the oscillator at the same time it adjusts the frequency-determining components of the mixer (in CB sets the

Communications Technology

channel selector changes the frequency-determining components). The arrangement is such that the oscillator will always be tuned 455 kHz above the frequency of the carrier being accepted by the mixer. The output of the oscillator is fed to the mixer.

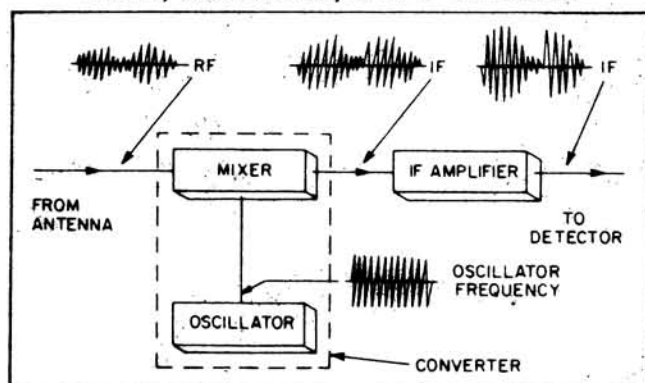
ANTENNA AND MIXER STAGE



In CB sets the frequency to which the oscillator is tuned is often two to 11 MHz above (or below) the incoming RF carrier frequency, instead of 455 kHz.

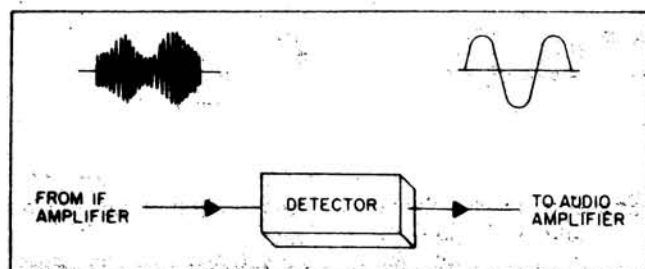
The Mixer—The carrier and oscillator frequencies combine in the mixer stage and four different frequencies appear at the output. One of these four is the difference between the oscillator and the carrier frequencies, and is usually 455 kilohertz. The other three are rejected by the next stage. The mixer and oscillator together are called the **converter**.

MIXER, OSCILLATOR, AND IF AMPLIFIER



The IF Amplifier—The abbreviation for **intermediate frequency** is **IF**. In most home receivers the IF is 455 kHz. Amplifying a single frequency in the IF circuit is much easier and causes less distortion than if it were necessary to tune this amplifier to each of the many station frequencies. The only purpose of this stage is to

DETECTOR STAGE

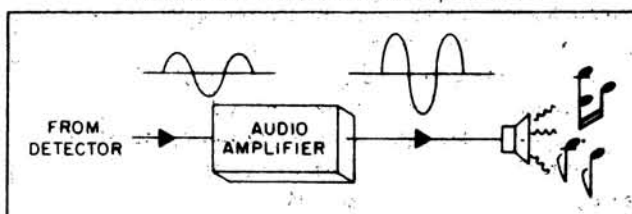


amplify the IF (which still retains the original audio frequency) and pass it on to the detector.

The Detector—The purpose of the detector is to remove the audio component from the IF waveform. The audio envelope is the same (although reversed) at the top of the waveform as it is at the bottom. The detector circuit is so designed that it accepts only the audio frequency at the top and rejects the IF frequency in the waveform.

The Audio Amplifier—The final circuit in the receiver amplifies the AF fed to it by the detector. The amount of amplification can be varied by the volume-control knob on the front of the receiver. The output of the audio amplifier is applied to the speaker voice coil, causing the speaker cone to reproduce the sound that originated at the studio.

AUDIO AMPLIFIER AND SPEAKER



QUESTIONS

- Q6. The _____ removes the AF from the IF waveform.
- Q7. The oscillator develops a signal at a constant _____ and _____.
- Q8. A converter combines the functions of _____ and _____.
- Q9. The IF (intermediate frequency) of most home AM sets is (455, 680, 1041) kHz. CB sets often have IF amplifiers set for (2-11, 5-20) MHz.

ANSWERS

- A6. The detector removes the AF from the IF waveform.
- A7. The oscillator develops a signal at a constant amplitude and frequency.
- A8. A converter combines the functions of mixer and oscillator.
- A9. The IF of most home AM sets is 455 kHz. CB sets often have IFs set at 2-11 MHz.

FREQUENCY MODULATION

The transmitter and receiver with which you have just become familiar employs amplitude modulation (AM) to carry the audio. Another method of superimposing audio on a carrier is called **frequency modulation** (FM). Its process is quite different. The two are compared in the drawing.

Both AM and FM start out with a carrier frequency and an audio frequency (sound originating in the studio). In amplitude modulation, as you already know, the sound is superimposed on the carrier frequency (which is constant) by varying the carrier **amplitude** in conformance with the voltage and frequency of the audio.

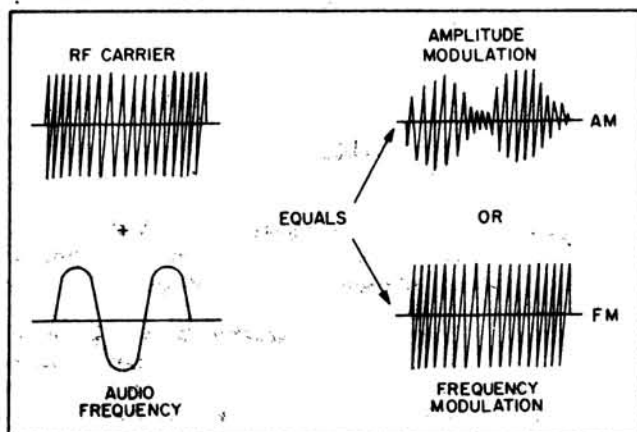
In FM, however, the audio is mixed with the RF in such a way that the carrier **frequency** is varied in accordance with the amplitude of the sound. As the

audio cycle goes positive, the RF carrier frequency increases. When the audio cycle goes negative, carrier frequency decreases. The sum of the two changed frequencies in one audio cycle is still equal to the original carrier frequency.

One of the advantages of frequency modulation is its freedom from distortion. Noise and other forms of distorting voltages in the atmosphere or receiver are added to amplitude modulation. Since FM does not depend on changing amplitude to carry audio, noise has little or no effect on it. This is part of the reason for the clarity of sound that you get from an AM receiver.

CB sets for many years (1958 until the early 70s) used AM to carry the audio information. Now, in order to get more distant reception, and to get more stations into a given amount of radio frequency spectrum, CB sets which use single sideband transmission are being manufactured. This is more complicated than ordinary AM, or even than FM. It is a special kind of AM, and will be used increasingly in the future, as more and more CB sets keep going on the air and crowding the radio spectrum.

COMPARISON OF AM AND FM WAVEFORMS



QUESTIONS

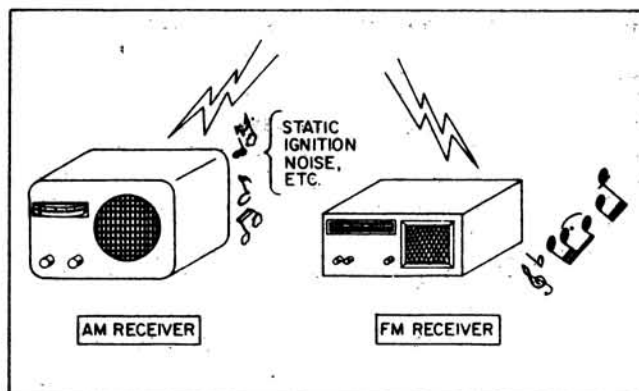
- Q10.** In AM, the carrier _____ changes to match the audio.
- Q11.** In FM, the carrier _____ changes to match the audio.
- Q12.** An FM receiver is (more, less) subject to atmospheric noise than an AM receiver.
- Q13.** Single sideband (SSB) transmission is a special kind of (AM, FM) which has some advantages over standard amplitude modulation.

ANSWERS

- A10.** In AM, the carrier **amplitude** changes to match the audio.

- A11.** In FM, the carrier **frequency** changes to match the audio.

NOISE HAS LITTLE EFFECT ON FM



- A12.** An FM receiver is **less** subject to atmospheric noise than an AM receiver.

- A13.** Single sideband (SSB) transmission is a special kind of **AM (amplitude modulation)** which has advantages over regular AM.

WHAT YOU HAVE LEARNED

All carrier signals within range are picked up by the receiver antenna. The tuning control on the front of the receiver adjusts the input of the mixer or converter stage so that only the desired station carrier frequency is received. At the same time, it adjusts an oscillator to generate an IF above the carrier frequency. Carrier and oscillator frequencies are joined in the mixer or converter and the difference between the two, the intermediate frequency, is amplified and fed to the IF amplifier. Here the signal and its audio component are further amplified. The next stage (detector) extracts the audio component and passes it to the final stage (audio amplifier). The audio is amplified and fed to the speaker, causing the cone to reproduce the sound that originated at the studio.

Amplitude (AM) and frequency (FM) modulation are two methods of transmitting audio on a carrier. When AM is used, the amplitude of the carrier varies according to the loudness (amplitude) and frequency of the audio. In FM, the frequency of the carrier is varied instead of the amplitude. FM transmissions are less bothered by atmospheric and receiver noises.

Single Sideband (SSB) is a special kind of AM which has the advantage of being received over greater distances than standard AM signals of the same power. It also uses up less space in the radio frequency spectrum, so a given amount of frequency spectrum can accommodate more transmitters.

Radio Transmissions

WHAT YOU WILL LEARN. When you have finished reading this article you will have learned what the electromagnetic frequency spectrum is, what a radio transmitter is, how it develops a broadcasting signal, and how

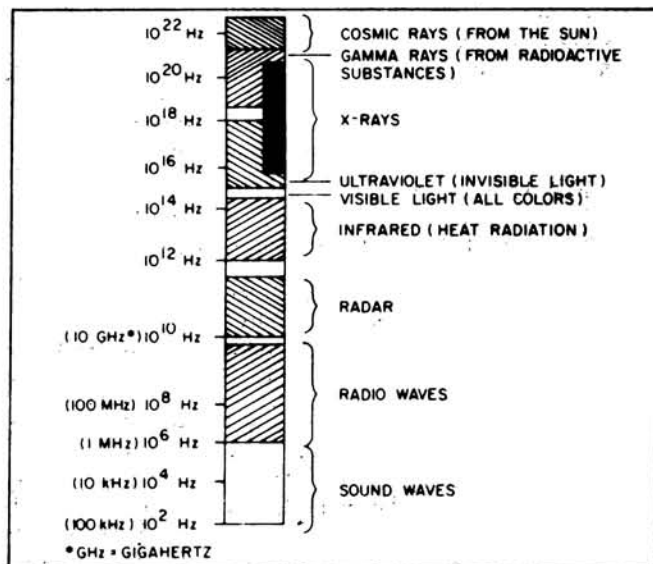
radio signals are transmitted through space. You will have learned that voice (or music) can be impressed on the radio (carrier) waves. In addition you will have become acquainted with the difference between amplitude modulation and frequency

modulation of radio frequency transmissions.

Electromagnetic Radiations

Energy that radiates from a source is said to be an **electromagnetic wave**. Gamma rays, which are given off by radioactive particles such as radium, uranium, or atomic-bomb fragments, are electromagnetic waves. **Cosmic rays** from the sun travel 93 million miles to the earth as electromagnetic waves. All electromagnetic waves, including light, radiated heat, and radio signals, travel through space at the rate of 186,000 miles per second.

THE ELECTROMAGNETIC SPECTRUM



The Electromagnetic Frequency Spectrum

Electromagnetic radiations differ from each other in terms of their frequencies measured in Hertz. (Hertz, a term we now use which means cycles-per-second). The frequency of one of these kinds of radiation is the number of times a single cycle repeats itself in one second. An electromagnetic spectrum chart shows the relationship of these different radiations to each other.

The chart shows that cosmic rays are radiated at a frequency of around 10^{22} Hz (abbreviation for Hertz). (The number 10^{22} is 1 followed by 22 zeroes, or ten-thousand, million, million, million Hz). At the lower end of the radio spectrum, the radiation frequency is under 10^4 , or ten thousand Hz.

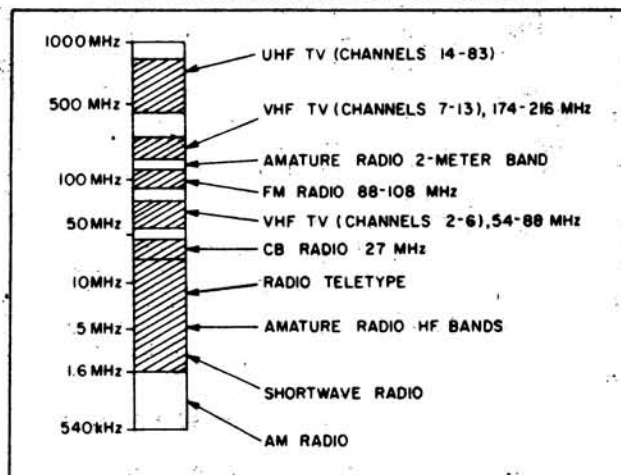
Assigned Broadcast Frequencies

The Federal Communications Commission (FCC) has assigned specific groups of frequencies to different types of communications transmissions. This is shown in an expansion of the radio-frequency portion of the spectrum in the right-hand chart.

Commercial transmitters (radio and television, for example) are assigned a transmitting frequency in the appropriate part of the radio-frequency spectrum. Transmitters broadcasting in the AM radio band, 535

kHz to 1,605 kHz, are required by law to be on their assigned frequency within plus or minus 20 Hz.

THE RADIO-FREQUENCY SPECTRUM



QUESTIONS

- Q1. Cosmic rays and radio waves are examples of
- Q2. Sound (is, is not) electromagnetic radiation.
- Q3. Radio waves travel from the broadcast station to a receiving antenna at the rate of miles per second.
- Q4. is the characteristic which distinguishes one electromagnetic wave from another.
- Q5. Commercial radio transmissions are at a (higher, lower) number of Hertz than television.
- Q6. A frequency of 1,000 kHz would be assigned to (commercial broadcasting, short-wave) radio.
- Q7. CB radio is located (above, below) 30 MHz.

ANSWERS

- A1. Cosmic rays and radio waves are examples of **electromagnetic radiations (waves)**.
- A2. Sound **is not** electromagnetic radiation. Remember? It is changing air pressure.
- A3. Radio waves travel from the broadcast station to the receiving antenna at the rate of **186,000 miles** per second.
- A4. **Frequency** is the characteristic which distinguishes one electromagnetic wave from another.
- A5. Commercial radio transmissions are at a **lower** number of Hz than television.
- A6. A frequency of 1,000 kHz would be assigned to **commercial broadcasting** radio. (1,000 kHz is equal to 1 mc.)
- A7. The CB radio channels are **below** 30 MHz (27 MHz).

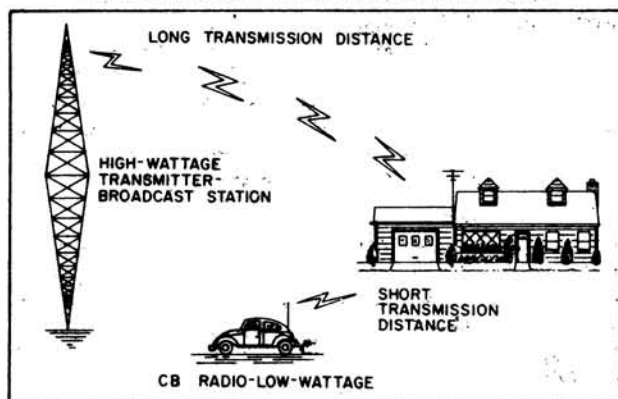
RADIO TRANSMITTERS

The dial on your home radio receiver is marked off in numbers, probably from 550 to 1,600 kHz (marked 55 to 160 often). By rotating the tuning dial, you select

the desired station. Since each local station broadcasts at a different frequency, you are able to select the one you desire. The dial setting indicates the carrier frequency of the station.

The channel numbers on CB transceivers (combined transmitter and receiver) may be any from "1" to "40." The channel number selects the exact frequency near 27 MHz. Channel 9, for example, is 27.065 MHz, and channel 11 is 27.085 MHz.

TRANSMISSION POWER AND DISTANCE



Transmitter Power

You have also noted that some stations come in stronger than others. The stronger stations broadcast at higher power (measured in **watts** or **kilowatts**) than the weaker. Or, if one of two stations broadcasting at equal power is stronger than the other, the stronger station is closer to your home.

The drawing shows two radio transmitters, one a powerful commercial broadcast station (AM broadcast stations generally transmit at least 250 watts, and no more than 50,000 watts), and the other a typical Citizens Band transmitter (limited to four watts). Even though the broadcast station is dozens, often hundreds of miles from the receiver, its signal is able to reach across that distance. The CB transmitter, on the other hand, will only go a few miles. Many broadcast stations can be received no farther than 20 or 30 miles away (sometimes less, due to obstructing hills or buildings).

Radio Frequency Carrier and Audio Frequencies

The frequency assigned to a broadcast station is called its **carrier frequency**. The transmitter and its antenna are designed and tuned to that specific frequency. As its name implies, the carrier frequency carries a reproduction of the sound originating in the studio. Actually, there are two frequencies that leave the transmitter, a **radio frequency** (carrier) and an **audio frequency** (sound). Audio frequencies are between 20 and 20,000 Hertz. The frequency range of most human ears, however, is usually no higher than 15,000 Hz.

QUESTIONS

- Q8. A home radio receiver (can, cannot) be tuned to 1 megahertz.
- Q9. 900,000 Hz (could, could not) be a carrier frequency of a commercial broadcast station.
- Q10. The power of one transmitter, station A, is 5,000 watts. Transmitter B broadcasts 500 watts. Which transmitter will be receivable at the greater distance?
- Q11. Two broadcast stations are equally distant from

your home. Assuming your receiver is OK, what could be a reason you could not receive one of them?

- Q12. A human ear (can, cannot) hear a radio frequency.
- Q13. A frequency of 600 kHz is classified as a (an) (audio, radio) frequency.
- Q14. CB radio transmitters (can, cannot) transmit up to 15 watts of output power.

ANSWERS

- A8. A home radio receiver can be tuned to 1 megahertz. One megahertz (1,000 Hz) is within the broadcast band.
- A9. 900,000 Hertz could be a carrier frequency of a commercial broadcast station. It is the same as 900 kHz.
- A10. Station A. It has much more power.
- A11. One station is so weak in power it cannot transmit the distance.
- A12. The human ear cannot hear a radio frequency.
- A13. A frequency of 600 kHz is classified as a radio frequency.
- A14. CB radio transmitters cannot transmit 15 watts output power (limit is 4 watts).

A Basic Transmitter

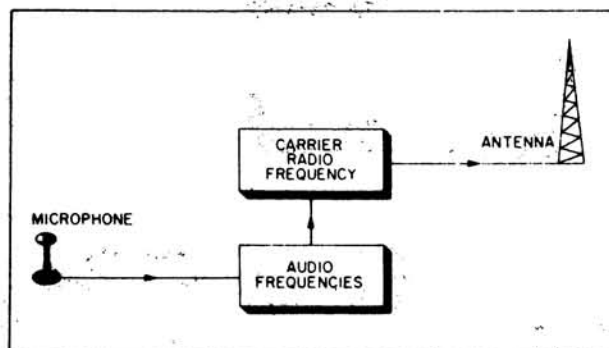
The diagram below shows a functional block diagram of a typical radio transmitter. It is called a functional block diagram because each block is representative of a general electronic function and may include several circuits.

The arrowheads between blocks show the direction of signal flow. You can probably already read what the diagram reveals.

Sound enters the **microphone** and is fed to the audio-frequency (AF) section. The sound, because it is too weak for transmission purposes, is amplified (signal amplitude is increased) and then passed to the RF carrier-frequency section.

Carrier Frequency—The specific radio frequency (RF) assigned to the transmitter is developed in the carrier-frequency block. Passing through several circuits, the RF signal is boosted in power (increased in amplitude) to the rated wattage output of the transmitter. Just before the RF carrier is fed to the antenna, the AF signal is superimposed on it. Waveforms developed in each block are shown below.

TRANSMITTER FUNCTIONAL BLOCK DIAGRAM



Superimposing the Sound—The process of superimposing audio on the carrier, as shown in this particular example, is called **amplitude modulation (AM)**. In amplitude modulation the audio frequency (varying

at the changing rate of the original sound) is mixed with the carrier (a constant frequency) in a manner that causes that carrier **amplitude** to vary at the same rate as the audio. The carrier **frequency** remains unchanged.

QUESTIONS

- Q16. The drawing here which shows circuit functions is called a (an) diagram.
- Q17. Sound enters the AF amplifier section from a device called a (an)
- Q18. on a block diagram show the signal direction between blocks.
- Q19. Placing an audio frequency on an RF (radio frequency) carrier without changing the frequency is called
- Q20. In CB radio the carrier frequencies are around Hertz. The audio frequencies are between (20 & 15,000, 50 & 5,000) Hertz.

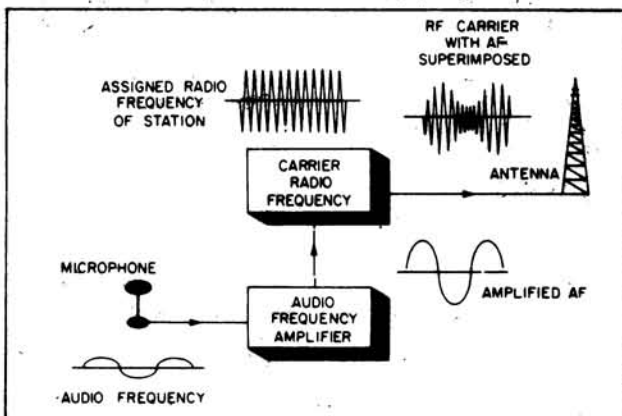
ANSWERS

- A16. The drawing is called a **functional block diagram**.
- A17. Sound enters the AF section by way of a device called a **microphone**.
- A18. **Arrowheads** on a block diagram show the signal direction between blocks.
- A19. Placing AF on a carrier without changing the carrier frequency is called **amplitude modulation**.
- A20. In CB radio the carrier frequencies are around **27 MHz**. The audio frequencies are typically between **50 & 5,000 Hertz**.

Carrier-Frequency Circuits

A minimum number of RF carrier-frequency circuits are shown in the diagram below. An actual transmitter has many more circuits to attain the frequency stability and power required. The additional circuits are similar to those shown, however.

TRANSMITTER WAVEFORMS



The Oscillator—The purpose of the **oscillator** is to generate a stable RF signal. The resistance, inductance, and capacitance that make up its input circuit are such that they will not allow the transistor in the oscillator to amplify any other signal but that of the desired frequency. The stable-frequency, low-amplitude output of the oscillator is shown above.

The Buffer—This stage (another name for circuit) is

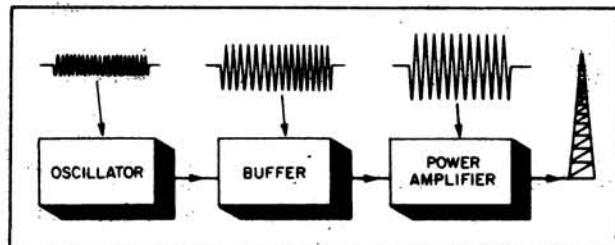
sometimes called an **intermediate power amplifier**, or **frequency multiplier**. In most transmitters it performs three functions. As a **buffer**, the stage isolates the oscillator from the effects of the circuits which follow it. Without this isolation, stray signals may be fed back to the oscillator, causing it to operate at the wrong frequency. As an **amplifier**, the buffer increases the amplitude of the oscillator signal to a level that is between the desired transmitter output and the amplitude of the oscillator signal. In many transmitters, the buffer circuit **doubles** (or even **triples**) the frequency of the oscillator output. The oscillator may not be capable of generating the required high frequency by itself. In order to produce the assigned frequency, a transmitter may require several multiplier stages.

The Power Amplifier—The purpose of the **power amplifier** is to increase the amplitude of the RF signal to the power (wattage) requirements of the station. Several stages of power amplification may be required to achieve this. Normally, the audio signal from the AF circuitry is fed to the final power amplifier and used to modulate the carrier.

QUESTIONS

- Q21. A transmitter circuit which amplifies a signal and increases its frequency is called a (an)

RADIO-FREQUENCY CIRCUITS



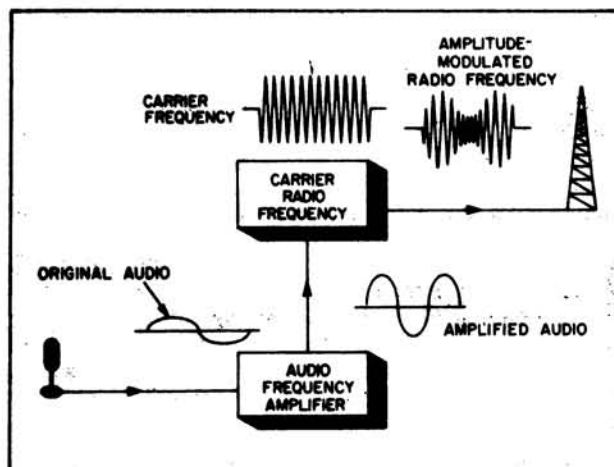
- Q22. A (an) generates a signal which has a uniform frequency.
- Q23. amplifier output is measured in watts.
- Q24. AF and RF are mixed in what stage?
- Q25. The carrier arrives at the antenna with its waveform (amplitude, frequency) modulated.
- Q26. If a CB transmitter, employed an oscillator which produced a basic radio frequency of 13.5 MHz, what kind of stage must be used after the oscillator to produce the required output radio frequency carrier of 27 MHz? (*tripier, doubler*).

ANSWERS

- A21. A transmitter circuit which amplifies a signal and increases its frequency is called a **multiplier**.
- A22. An **oscillator** generates a signal which has a uniform frequency.
- A23. **Power** amplifier output is measured in watts.
- A24. AF and RF are mixed in the **final stage of the power amplifier**.
- A25. The carrier arrives at the antenna with its waveform **amplitude** modulated.
- A26. To produce a CB radio frequency carrier of 27 MHz a 13.5 MHz oscillator would have to be followed by a **doubler** stage.

Audio-Frequency Circuits

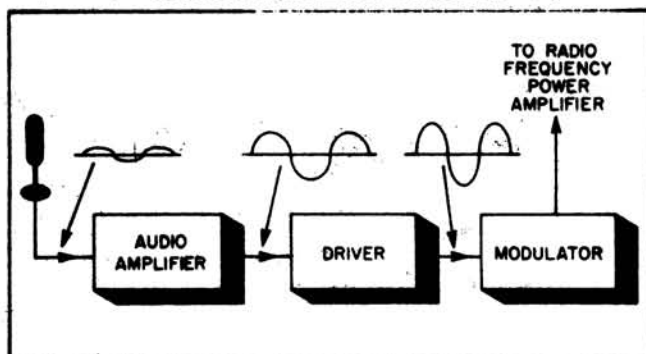
MIXING AUDIO AND RADIO FREQUENCIES



The Microphone—Regardless of the many different types of microphones that are available, even the best develop only a weak signal.

The Audio Amplifier—Although a single stage of audio amplification is sometimes all that is necessary, larger transmitters may have two, three, or more stages to obtain the desired undistorted level of amplitude.

AUDIO-FREQUENCY CIRCUITS



The Driver—Like most circuits, the driver obtains its name from its purpose. The driver amplifies the AF to the voltage level required to "drive" the transistors of the modulator. The modulator transistor requires large changes in signal amplitude to operate properly.

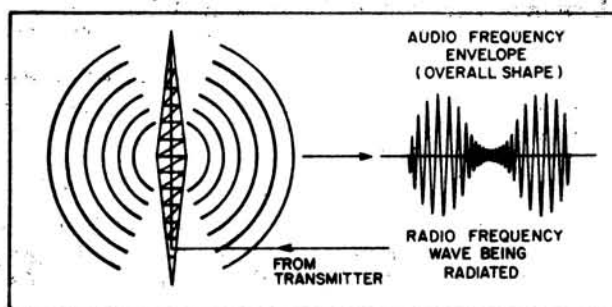
The Modulator—The modulator is a power amplifier quite similar to the final circuit of the carrier-frequency block. It amplifies the audio signal to a power level suitable for modulating (changing) the carrier power in the final power amplifier. Power output of the modulator is fairly close to half the power of the final carrier amplifier.

Antennas

If all circuits are operating properly, an AM (amplitude-modulated) carrier is fed to the antenna.

Power is fed to the antenna in the form of both current and voltage. Voltage sets up an electric field along the length of the antenna. Current, in traveling through the antenna (a conductor), sets up a corresponding magnetic field. Both fields vary at the rate of the carrier frequency and at the amplitude and frequency of its audio envelope.

ANTENNA RADIATION



Both fields expand outward and collapse back to the antenna at the rate of the carrier frequency. The outermost waves continue through space and do not return to the antenna. This action is similar to dropping a pebble in a pool. The energy of the waves moves outward in ever-widening circles; the water, however, remains in place.

QUESTIONS

- Q27. The weak output of a microphone is fed to one or more stages of amplification.
- Q28. The output of even the best microphones (can, cannot) be fed directly to the modulator.
- Q29. The output of the is connected to the carrier power amplifier.
- Q30. For proper modulation, the output of the modulator stage must be that of the power amplifier.
- Q31. Carrier voltage develops a (an) field and carrier current develops a (an) field on the antenna.
- Q32. All of the energy in the antenna fields (does, does not) leave the antenna.

ANSWERS

- A27. The weak output of a microphone is fed to one or more stages of **audio** amplification.
- A28. The output of even the best microphones **cannot** be fed directly to the modulator. (Even the most powerful microphones develop a signal that is much too weak to drive the modulator.)
- A29. The output of the **modulator** is connected to the carrier power amplifier.
- A30. For proper modulation, the output of the modulator stage must be **half** that of the power amplifier.
- A31. Carrier voltage develops an **electric** field and carrier current develops a **magnetic** field on the antenna.
- A32. All of the energy in the antenna fields **does not** leave the antenna. (Only the outermost waves.)

WHAT YOU HAVE LEARNED

1. Radiant energy is given off by electromagnetic waves. The electromagnetic spectrum includes cosmic rays, X rays, visible and invisible light, infrared, radar, as well as radio waves.
2. A radio transmitter is a device that produces electromagnetic waves in the radio portion of the spectrum. Its essential functions are the development and amplification of a carrier frequency and modulating it with an amplified audio frequency. A specific carrier frequency is assigned to each radio station. The distance that the carrier, with its super-

imposed audio, travels is determined by the power that is developed in the final stage.

3. Energy in the form of voltage and current is fed from the transmitter to an antenna. This sets up electric and magnetic fields around the antenna

that expand and collapse at the frequency of the carrier. Part of the energy is in the form of electromagnetic radiations and is transmitted through the atmosphere. The farther it travels, the weaker the signal becomes. ■

Signal Propagation

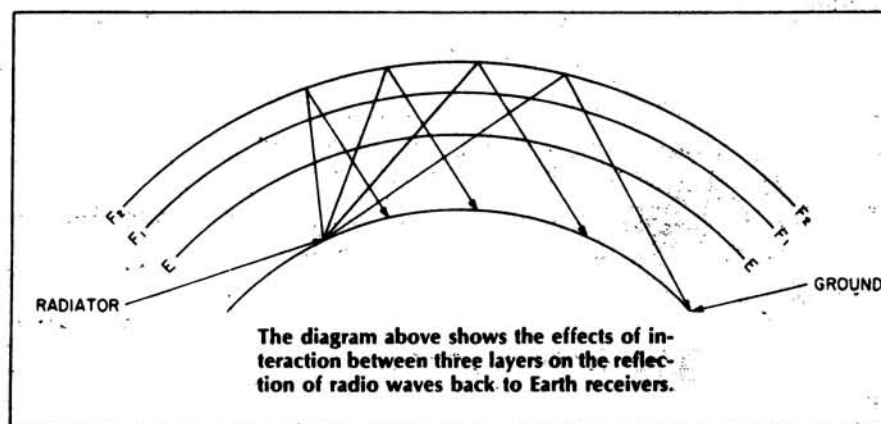
□ The propagation of all radio waves may be defined as: "The traveling of wave energy through the atmosphere." How do signals get from one side of the world to the other? The ionosphere, a region of the earth's atmosphere extending upwards from 50 to 350 miles (80-560 kilometers), has ionized layers with the ability to reflect electromagnetic energy of certain frequencies.

The purpose of this article is to review, briefly, the components of the ionosphere, some propagation characteristics, and some ways that you can tell whether or not the amateur bands are in good condition for DX communications.

The Ionosphere. In 1901, Marconi succeed in transmitting radio signals from England to America, over the bulge of the earth. Since it was known that electromagnetic waves travel in straight lines, Heaviside and Kennelly proposed in 1902 that the radio waves had been reflected from an atmospheric layer consisting of free electric charges. Heaviside called it the "electrified layer", and others called it the "Heaviside layer." Later, Appleton discovered another higher layer. To distinguish between the two layers, and to make room for other possible layers, he called the Heaviside layer the E layer, and the new layer the F layer. Subsequently a lower layer was discovered, called the D layer. It was also found that the F layer consisted of two layers which separated during the day and merged at night, and the concentration of electrons decreased in all layers at night.

The D-layer (50 miles or 80 km up) assembles during the day and disappears at night. It primarily affects 160 and 80 meters, limiting daytime coverage to groundwave by absorbing skywave signals.

The E-layer (75 or 120 km up) provides stable skywave coverage during the day on 10 through 40 meters, and on 80 and 160 at night. The nominal skip distance is around 1,000 miles (1,600 km), but double hops can and



do occur. During primarily the summer months, sporadic-E—extra-thick patchy clouds of ionization—opens 6 and 10 meters for short duration openings during the day or night.

The F-layer at night is about 200 miles (320 km) high, and during the day, it separates into the F₁-layer (125 miles or 200 km up), and the F₂-layer (250 miles or 400 km up). These layers are the mainstays of the DX'ers looking for worldwide coverage. A single hop covers 2,000 to 3,000 miles (3,200-4,000 km), and multiple hops cover up to 12,000 miles (19,300 km). What gets reflected by the F₁ and the F₂-layer depends upon the degree of ionization and the angle of the arriving signal at the ion layer.

The Maximum Useable Frequency. The maximum useable frequency (MUF) is simply defined as the highest frequency that can be used over a particular transmission path.

The MUF is affected by three variables: the 11-year sunspot cycle, the annual cycle, and the daily cycle.

Mid-1980 is generally predicted as the next peak in the sunspot cycle. Although the upcoming peak is not expected to equal the all-time historical cycle centering on 1958 when the MUF cleared 100 MHz on many occasions, 10 and 15 meters should be superlative.

The annual cycle varies from summer to winter. Due to the amount of

solar radiation received by the several ion layers, the MUF is lower during the winter than summer (although atmospheric noise during the summer may hide those benefits). The overall height of the F₂-layer during the winter day is lower too, by some 200 miles (320 km) than during the summer day. The F₁-layer is also lowered somewhat. The effect is that more hops are required to cover the same distance, thus introducing more variables into the degradation of the signal over the transmission path. Winter or summer, the height of the nighttime F layer remains about the same.

The daily cycle is the most variable. The MUF is higher during the day than at night. During the short winter day, the "day" effect may seem brief. The twilight hours of DX'ing during the winter may be longer, and very interesting.

Twilight DX'ing. The rate of change of ionization is greatest at sunrise and sunset. (Listen to the broadcast band (540-1600 kHz) for a vivid example of this.)

80 and 40 meters provide very interesting DX'ing conditions worldwide during the twilight hours. The specialty of greyline DX'ing, sending signals around the world following the day-night demarcation, is a science in itself. The key is in timing, knowing when both ends of the path will be in twilight.

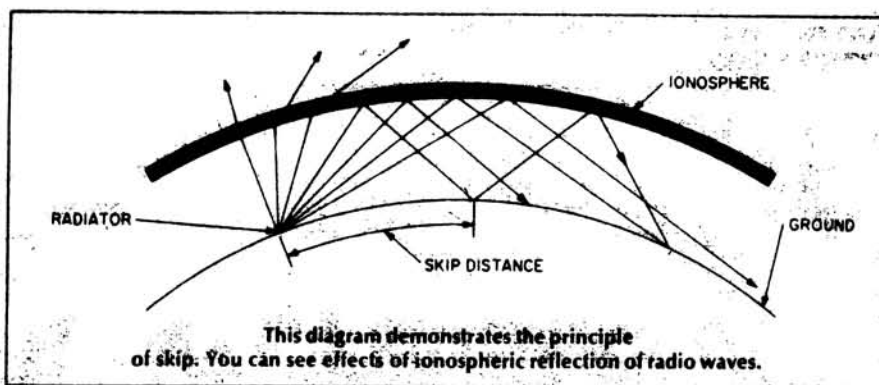
On the higher bands, DX'ing the twilight hours, by pointing your antenna into the daylight, can maximize the coverage of your signal. In the daylight, the ionization at the reflection points of the transmission path are near maximum, and any D-layer absorption is rapidly disappearing.

Band Conditions. Aside from the three-fold variables affecting the MUF, there are other short-term, solar-induced factors that affect propagation.

Sunspots, storms on the surface of the sun, can affect conditions by several types of radiation. Sudden ionospheric disturbances (SIDs) caused by solar flares can cause instant blackouts of worldwide radio communications on the earth.

Other, less severe, disturbances caused by intense radiation from the sunspot (or group of sunspots) tend to be longer lasting. As the radiation is sprayed outwards (just like pointing a garden hose away from you and turning quickly in a circle), sweeping across the earth, the effects on the earth may last from four to five days at a time. The sun rotates on an average 27 period, and the long-lived sunspot eruptions may have a recurring effect on the earth approximately 27 days, and multiples thereafter, until the particular sunspot(s) disappears.

The Auroral Zones. The magnetic poles of the earth draw most of the particles toward the north and south polar regions. These auroral zones are



centered roughly 23 degrees from the magnetic poles.

The size of the auroral zones changes with the intensity of the solar radiation. Flares and sunspots will increase the geomagnetic activity in the polar regions.

Signals traversing the auroral zones will have a characteristic polar flutter induced on them. If the geomagnetic storm is severe enough, signals will be absorbed.

As signals travel a Great Circle path, the high latitude paths will be the first to disappear as these reflection points are absorbed. As the storm continues, the lower latitude signal paths will disappear into the noise and, eventually, only signals from the south (in the northern hemisphere) will be

heard. Those reflection points are outside the enlarged zone.

Sources of Information. How do you know if band conditions are good enough to warrant your time and effort to chase some DX? One time-consuming way is to tune across the bands, stopping at each signal encountered to see who it is.

The most timely information is provided by WWV. Unlike the other sources cited above, the National Bureau of Standard's propagation information is oriented for the engineer and scientist, and not for just the radio amateur.

WWV's Propagation Bulletin. At 18 minutes past each hour, WWV (on 2.5, 5, 10 and 15 MHz) broadcasts a tape supplied by the National Oceanic and Atmospheric Administration's Environmental Research Laboratories.

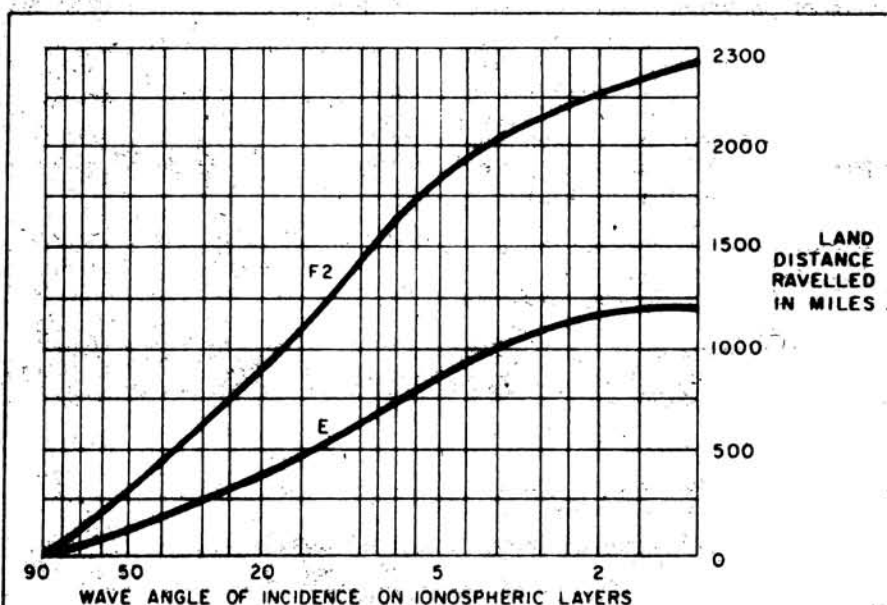
The message contains number counts of the solar flux, the A-index, and the K-index.

The solar flux is a measure of solar radiation; units range from 60 to 400. The higher the flux, the higher the MUF. In late 1978, the flux was averaging 160.

The A-index is a measure of geomagnetic activity ranging from 0 (extremely quiet) to 400 (very disturbed) units. An A-index of more than 20 usually wipes out polar path signals. Only on a very few occasions in the past five years have I seen this index exceed 100.

The K-index is mathematically related to the A-index and is a measure of geomagnetic activity. Its range is 0 to 9.

As the flux and A-index counts are updated daily at 0000 GMT, and the K-index is updated every six hours at 0000, 0600, 1200, and 1800 GMT, it is possible to track short-run variations from day-to-day and from six-hour



The graph above shows the relationship between the angle of incidence of radio waves striking the "E" and "F₂" layers of the ionosphere and the distance which they will travel over land. Obviously, the lower the angle, the greater the distance travelled by the waves.

period to six-hour period.

In practical terms, you can build a set of references for yourself by noting the WWV counts under normal and abnormal conditions. The set of references may differ for different opera-

tors. The better the equipment and the antennas, the longer you have to work DX under worsening or marginal conditions.

Conclusion. I have only touched on some the principles and highlights of

propagation and determining band conditions.

Amateur band DX'ing doesn't *have* to be hit or miss. Do a little reading and a little studying, and watch your DX totals go up. ■

AM/FM Receiver Alignment

□ The words "receiver alignment" often conjure up a mysterious and complicated procedure which can be performed only by an expert. While it is true that receiver alignment should not be attempted by anyone who does not have the necessary skills and equipment, it is not a very difficult procedure if you have some basic guidelines. This article will explain the various procedures for aligning AM, FM, and AM/FM receivers using a minimum of equipment.

Before getting into the mechanics of receiver alignment, it is important that the service technicians understand the basic operation of the receiver, and why proper alignment is necessary. To this end, a discussion of the modern superheterodyne receiver will follow.

Receiver Fundamentals. Virtually every AM and FM receiver manufactured today is a superheterodyne receiver. They utilize a built-in RF oscillator and mixer circuit to convert the received signal to a lower frequency called an intermediate frequency (IF). In a given AM or FM receiver, the IF remains the same frequency regardless of which radio station is being received.

The basic advantage of this type of circuit is that the greatest part of the RF gain, and bandpass characteristic of the receiver, is provided by the IF stages, and is a constant for all received frequencies. Thus, the manufacturer of the receiver can precisely determine the sensitivity and selectivity of the receiver. Proper alignment ensures that the receiver performs exactly the way the manufacturer intended. Refer to Fig. 1, which is a simplified schematic diagram of the RF and IF sections of a typical modern day AM receiver.

In Fig. 1, simplified AM radio schematic, you will note a combination

mixer-oscillator stage, and two stages of IF amplification. This circuit is typical of a minimum cost AM receiver which has a two gang tuning capacitor and no RF amplifier. Such a circuit might be used in the common pocket-sized AM transistor radio. More expensive receivers will use a similar circuit with the addition of a transistor stage for RF amplification.

In the circuit of Fig. 1, the received signal is picked up by the loopstick antenna and is fed to the base of Q1. This transistor stage is used as a combination local oscillator and mixer, and is actually a Hartley oscillator with the received signal being placed in series with the base drive of the oscillator. The frequency of oscillation is determined by C3, C4, and T1. At the same time C1, C2 and the loopstick antenna are tuned to the received signal frequency. The output of T2 contains several frequencies: the radio station frequency, the local oscillator frequency and two new frequencies equal to the sum and difference of the local oscillator and received frequency.

It is the function of the two IF stages, Q2 and Q3, to amplify the difference frequency (IF) and reject all others. This is accomplished by tuning all IF transformers, T2, T3, and T4, to the specified frequency. For most AM receivers, this is 455 KHz. The tuning of these transformers is performed during alignment of the receiver.

FM Circuitry. Fig. 2 is a simplified schematic of a typical FM receiver. In this diagram you will note the similarity with the AM receiver schematic of Fig. 1. This basic difference, aside from the higher operating frequency, is that the FM receiver employs an RF amplifier stage, Q1.

Most FM receivers have at least one

RF amplifier, since it is the nature of FM transmission that weaker signals than AM are usually encountered. Good FM reception requires a solid signal in the IF amplifier, so that effective limiting takes place. Note also the requirement of a three gang variable capacitor instead of the two gang as appears in Fig. 1. The additional cost and size of this capacitor is one reason why most AM receivers have no RF amplifier stage.

Although the IF amplifier stages of the AM receiver and FM receiver appear to be similar, there is a substantial difference in the way in which they are designed. The bandwidth of an AM radio station is just 10 KHz, and the receiver must be designed to have an IF bandwidth no greater than this. Such a narrow bandwidth is easily controlled by the design of the IF transformers which, when tuned to the same frequency such as 455 KHz, will provide the desired bandwidth of 10 KHz.

In the case of FM reception, and especially Stereo Multiplex FM, good reception requires a receiver with at least 150 KHz bandwidth. It is the nature of FM to have significant sidebands on either side of the carrier frequency, as far away as 75 KHz. Such a wide bandwidth in the IF stages of an FM receiver cannot be attained by tuning each stage to the same frequency. What is done is to stagger tune each stage, so that the resultant overall response has the required bandwidth. Note that each IF transformer in Fig. 2 has separate tuning slugs for primary and secondary which permit stagger tuned alignment.

Because of the problem of attaining proper alignment in the IF stages of an FM receiver, the more expensive designs utilize special filter circuits which

are tuned at the factory and require no adjustments in the field. You will find that many modern stereo receivers are designed this way. However, these types of receiver still require alignment of the RF and local oscillator sections of the unit as you would expect.

AM Receiver Alignment. Alignment of an AM receiver is a relatively simple procedure, and can actually be performed by using existing radio stations as a signal source instead of a signal generator. However, if a signal generator is available, it is always best to use it as described in the following paragraphs.

When performing the alignment of an AM receiver, connect a few loops of wire across the output cable of the signal generator and loosely couple the loops to the AM antenna coil. Use the smallest RF output from the generator that will produce a reliable meter reading. For a tuning indicator you can connect a VTVM or VOM set to measure AC volts across the voice coil of the speaker.

An alternate method is to measure the DC voltage level on the AVC (Automatic Volume Control) line of the receiver. This requires a high impedance DC voltmeter such as a VTVM.

The alignment procedure will, of course, depend upon the number and type of adjustments in the specific receiver under test. If possible, obtain the manufacturer's alignment procedure. In the absence of such information you can use the following procedure.

IF Alignment. The first adjustment to be performed is the IF alignment. Set the signal generator at 455 KHz. with about 30% to 90% amplitude modulation. Loosely couple the output of the generator to the antenna coil of the receiver and listen for the audio modulation in the receiver speaker as the RF generator is varied about 455 KHz. If you do not obtain any audio around this frequency, the IF of the receiver may be another frequency, such as 262 KHz.

Once you have determined the proper frequency, set the generator at 455 KHz. or whatever the specified intermediate frequency may be. Connect the VOM across the voice coil of the receiver and adjust the IF transformers (T2, T3, and T4 in Fig. 1) for the maximum reading of the meter.

As you progress with the IF alignment, you may find that you can lower the RF output of the signal generator to prevent overload of the receiver. When you are satisfied that all IF transformers are tuned so that no further

increase in meter reading can be attained, the IF alignment is complete.

The next set of adjustments will align the upper and lower settings of the tuning capacitor. This is accomplished in two steps, which will have to be repeated several times, since the adjustments have some interaction with each other.

The initial adjustments are made at

the high end of the broadcast band. Set the tuning dial to a silent point (where no radio station is received) near 1600 KHz. Set the RF signal generator to the same frequency. Adjust the two trimmer capacitors (C2 and C4 in Fig. 1 for maximum meter reading.

Then set the tuning dial of the receiver to a silent point around 600 KHz. and set the signal generator to

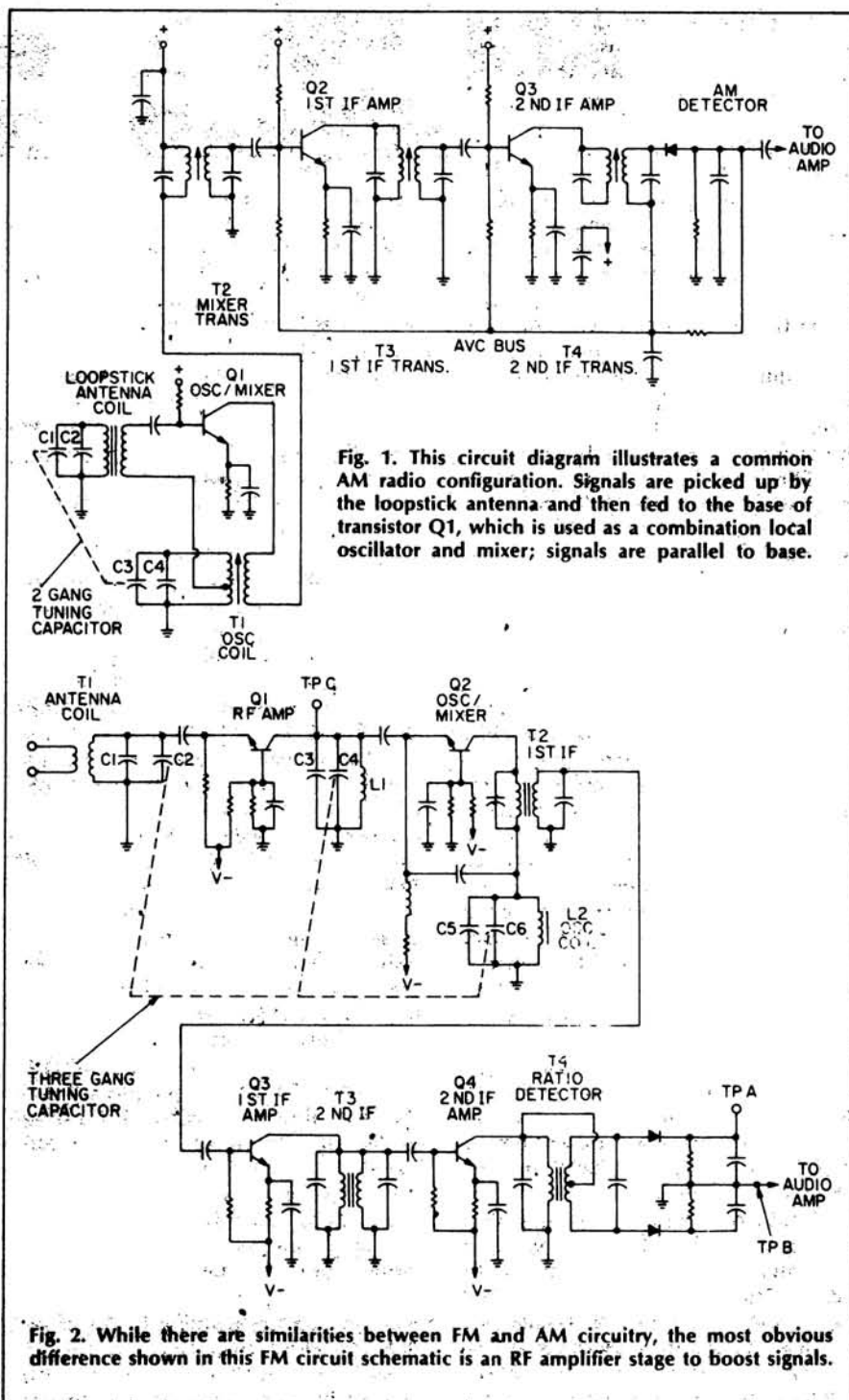


Fig. 1. This circuit diagram illustrates a common AM radio configuration. Signals are picked up by the loopstick antenna and then fed to the base of transistor Q1, which is used as a combination local oscillator and mixer; signals are parallel to base.

Fig. 2. While there are similarities between FM and AM circuitry, the most obvious difference shown in this FM circuit schematic is an RF amplifier stage to boost signals.

the same frequency. Adjust the oscillator tuning coil (T1 in Fig. 1) for maximum meter reading.

Repeat the adjustments for 1600 KHz and 600 KHz as described above until no further improvement can be made. Once you have done this, the AM alignment will be completed.

FM Receiver Alignment. Proper FM alignment requires the use of a sweep generator and oscilloscope, unless the receiver under test has a fixed tuned IF section which is permanently aligned at the factory. The generator may be coupled to the antenna terminals of the receiver using the circuit of Fig. 3, which matches the 50-ohm impedance of the generator to the 300-ohm input impedance of the FM receiver.

Some receivers may have highly selective front ends which make injection of the 10.7 MHz IF through the RF stages difficult. In these receivers you may couple the output of the signal generator to the input of the first IF amplifier stage through a 10-pf capacitor. This point is shown in Fig. 2 as test point C. Use only sufficient signal strength to achieve a usable display on the oscilloscope. Fig. 4 is a typical connection diagram of the receiver, generator and oscilloscope.

The first section of the FM receiver to be aligned is the IF amplifier. Set the signal generator to a center frequency of 10.7 MHz and a sweep width of about 450 KHz. An FM sweep generator should have a marker system which allows you to identify the important frequencies of interest, such as 10.6, 10.7, and 10.8 MHz. These markers indicate the desired bandwidth of the IF amplifier.

Connect the Y input of the oscilloscope to test point A of Fig. 2, and adjust the scope controls to obtain a display similar to Fig. 5. Adjust all IF transformers (T2, T3, & T4 in Fig. 2) so that the amplitude of the display is maximum, while preserving the symmetry of the waveshape at about 10.7 MHz. Bear in mind that low cost FM receivers will have a waveshape similar to that of waveform A, while high quality receivers will have the preferred waveshape B of Fig. 5.

The next step is the adjustment of the Ratio Detector or Discriminator. Connect the Y input of the oscilloscope to test point B of Fig. 2, and adjust the secondary of the Ratio Detector or Discriminator transformer (T4 of Fig. 2) to place the 10.7 MHz marker at the

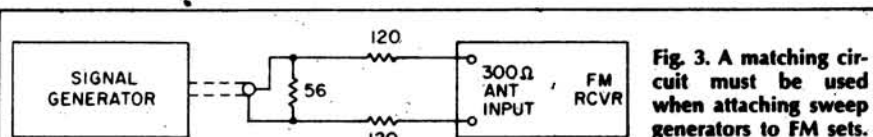


Fig. 3. A matching circuit must be used when attaching sweep generators to FM sets.

Fig. 4. When using both a sweep generator and oscilloscope for aligning an FM set, they must be connected in the manner shown in the diagram on the right side.

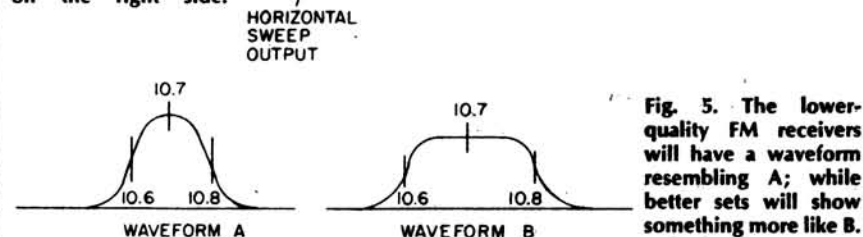


Fig. 5. The lower-quality FM receivers will have a waveshape resembling A; while better sets will show something more like B.

center of the S curve, as shown in Fig. 6. Readjust the primary of the same transformer for maximum amplitude and straightness of the line between the positive and negative peaks. This completes the alignment of the IF section of the receiver.

Trimmer Adjustments. The final adjustments to the FM receiver are made to the trimmer capacitors, which are connected in parallel with the main FM variable capacitors. These are shown in Fig. 2 as C1, C3 and C5. Note that some FM receivers, especially the higher quality ones, may have four such trimmer capacitors. These adjustments are made with the FM tuning dial set to near the high end of the range, (108 MHz).

Set the FM dial to a silent point near 108 MHz. Set the signal generator to the same frequency as the dial setting, and couple the output of the generator to the antenna terminals of the receiver as shown in Fig. 3. Connect a DC voltmeter to test point A as shown in Fig. 2, and use only enough signal output from the generator to attain a reliable meter reading. Adjust C1, C3, and C5 as shown in Fig. 2 for maximum meter reading.

As a final check, set the receiver dial to a silent point near the low end of

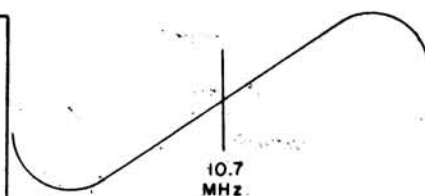


Fig. 6. Adjust the secondary of the discriminator transformer 'til it centers 10.7 MHz.

the range, 88 MHz, and set the generator frequency to the same frequency. Now, by varying the generator about 88 MHz while watching the DC voltmeter, you can determine how accurate the FM dial is. If necessary, you may adjust the oscillator coil L2, or slip the dial pointer so that the dial reading agrees with the signal generator frequency.

Keep in mind that if you do make this adjustment you will have to go back to 108 MHz and readjust the trimmer capacitors again. In addition, if you are working on an AM/FM receiver slipping the dial pointer will also require a realignment of the RF section of the AM receiver. For these reasons it is not recommended that the dial be slipped unless it already has been placed far out of calibration through some previous servicing procedure.

Satellite Receivers

□ In slightly more than 20 years, the nations of the earth have launched over 11,000 orbiting satellites! Nearly all of them are either American or Russian. Fewer than 200 have been launched by all other nations combined. Of course, many of those satellites have decayed from orbit, spending their last few seconds of life as a flaming fireball. At present, nearly 4700 objects remain in orbit.

Why put those satellites up there at all? Aren't they extremely expensive? And can't we perform adequate scientific experiments here on Earth without going into orbit?

Reliability of satellites is much better than that of shortwave radio. Normal radio communications links are highly

dependent upon the sun, often suffering massive blackouts during solar storms. Satellites do not have this vulnerability.

Nations of the earth have decided that satellites do offer an advantage over earth-bound technology. So far as the cost, the use of satellites is shared by many supporters, justifying the high initial costs. That distributed cost will be decreased even further with the implementation of the Space Shuttle. The shuttlecraft will be able to place satellites in orbit and even extract them for repair and subsequent replacement.

Ears and Eyes Aloft. You have been enjoying the benefits of earth satellites for many years. Weather satellites have been providing closeups of storm cen-

ters and weather system movements. Military satellites constantly comb the earth for signs of possible aggression. Domestic communications satellites have relayed your long-distance telephone calls. But the most obvious of the present generation of earth satellites are the television relay satellites.

At this writing, five primary TV relay satellites are now positioned over the equator where they favor North American viewing. Four are American and one is Canadian. They are: RCA SATCOM I, 24 channels, programmed primarily for cable TV users; RCA SATCOM II, 12 channels, supplies special programming contracted by networks; WESTERN UNION WESTAR I, supplies Public Broadcasting Network; WESTERN UNION WESTAR II, Spanish Information Network and non-network "superstations;" CANADIAN ANIK III, Canadian Broadcasting Company and French TV programming.

Other satellites are nearby, on standby status for backup. The transmitters on board the domestic satellites are called transponders. They act as repeaters for incoming signals on the "uplink" from Earth.

Prime Time Fare? Assuming that you have your terminal, what is there to see? With the satellites presently providing 12 or 24 channels of programming, the variety is incredible! Home Box Office with its first-run movies, Warner Star Channel with movies for all tastes, sports specials, cultural programming, religion, news; an endless variety of choices.

Tuning In. It would be nice if we could point a conventional TV antenna at one of them and tune it in. Unfortunately, satisfactory satellite TV reception is incredibly complex, requiring sophisticated receiving equipment. These satellites are not compatible with our TV sets. Some of the differences between normal home TV reception and satellite reception include: frequency band, signal strength, and modulation.

In order to conserve spectrum space, many domestic satellites employ "frequency re-use;" that is, 12 channels are transmitted into a horizontally-polarized antenna, and then 12 more channels are transmitted on the same frequencies using a vertically-polarized antenna. Thus, the receiving user must match the antenna polarization to receive the signals he wants to see and hear.



Our home TV sets operate on VHF (54-216 MHz) and UHF (470-890 MHz). American TV satellites transmit on the C-band (3700-4000 MHz). And to make matters worse, the orbiting transmitters are 22,300 miles away! Little wonder that there are only a few hundred operational, home-assembled satellite TV receiving terminals.

All of the domestic satellites are positioned in a straight line above the equator. Because they move around the earth at exactly the same rate as the earth rotates on its own axis, these satellites are called "geosynchronous." Thus, because they appear to remain permanently positioned overhead, they are called "geostationary." Once your antenna is pointed at a satellite, further tracking is unnecessary.

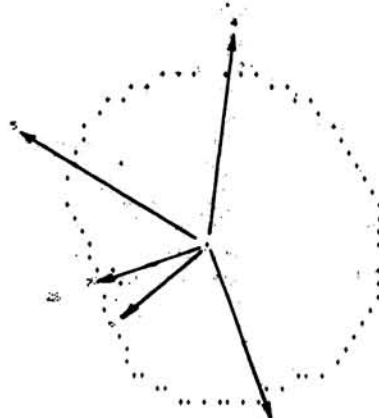
It is a ticklish business keeping them separated far enough to avoid interference on common channels (detected as "sidelobes" off the edges of directional antennas), while keeping them close enough for adequate signal strength for American users. The TV satellites are positioned above the equator from approximately 99° to 135° west longitude. Their signal radiation patterns (called "footprints") are established to favor the population densities of the north central United States.

After the signal is located by the giant antenna (usually a parabolic dish of at least 10-feet in diameter), it must be amplified by a low-noise amplifier (LNA). Amplifiers like this are hard to come by, costing at this writing nearly \$1000. The special low-noise transistors (and there are usually two in each LNA) may cost over \$200 apiece! Because of their special internal composition of Gallium Arsenide, these field effect transistors are called "GASFETS" (GaAsFETs). The reason for such

ES IDENTIFIER	LINK	MA	71	18	20
1 BEAR HILL	158 1	7 2	0	0	0
2 ASHUMHUS IT	252 4	51 5	19	6	12
3 BEAR HILL	158 1	7 2	7	0	7
4 CHESILE	6 5	54 0	12	17	0
5 TROY	201 4	86 9	18	18	0
6 NOBSCOT	220 5	16 2	1	0	1
7 ASHUMHUS IT	252 4	51 5	7	5	2
8 ASHUMHUS IT	252 4	51 5	16	6	10

PROBABILITY OF COORDINATION = 47 80 (PARABOLIC) FAIR
86 50 (HORN) EXCELLENT

This is a sample of what the microwave survey provided by Channel One, Inc. looks like.



meticulous choice of parts is simple. Most amplifying devices work at low frequencies, and at reasonable signal levels. The weak signals from the TV satellites are nearly indistinguishable from cosmic background noise. Conventional amplifying devices become worse—generate their own internal noise—as their operating frequencies are raised. More noise we *don't* need! Consequently, acceptable transistors must be carefully manufactured and hand picked. Even the LNA circuitry itself is unconventional—stripline tuned circuits. Definitely not the sort of project for the average weekend tinkerer!

Even after the satellite signal has

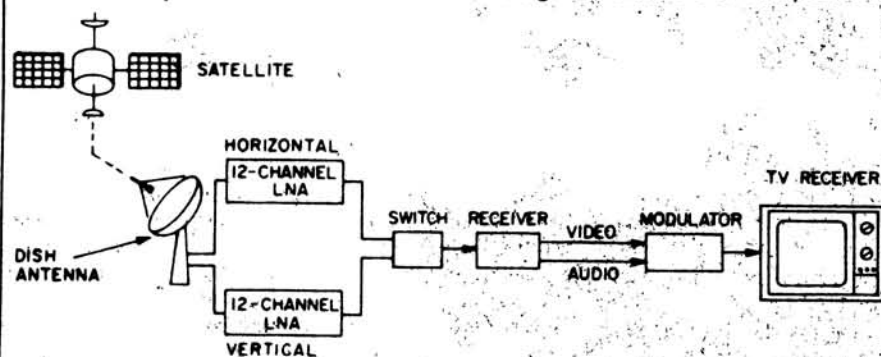
been amplified, there is need to process it further. A satellite receiver is used to provide video and sound to a TV monitor and speaker. Your own TV set cannot be used directly without a base-band demodulator—a device similar to that found in add-on TV games which may be connected directly to the antenna terminals of your TV set. Thus, the 4 GHz satellite signal is down-converted to channel 3 or 4 so that you may have the economical convenience of using your own TV set to watch the satellites directly.

Rolling One's Own. Can inveterate experimenters piecemeal together a working satellite terminal? Although some technical types have successfully converted surplus microwave receiving equipment into TV satellite receivers, the process is very difficult. It should not be considered by most hobbyists.

Unquestionably, prices will drop dramatically in the near future. But for now, consumer demand is not great enough to justify mass production techniques, so prices remain inordinately high. How high? Several thousand dollars for the least expensive terminal.

Buying In. As with any new technology, a large number of manufacturers are appearing on the market. It is still too early to tell which of these companies will grow to dominate the field of home TV terminals, but several companies were particularly helpful in pro-

This is a simplified block diagram of what a typical home-use earth receiving station would look like. The 5-meter dish antenna in your yard feeds two signal amplifiers; one for the vertically polarized signals, and one for the horizontally polarized ones. A switch in the receiver selects either amplifier's output to be demodulated, and then fed to a special modulator which converts the signal to channel 3 or 4 on your TV.



viding information in the preparation of this article. We are pleased to mention the availability of their services.

Spacecoast Research (P.O. Box 442, Altamonte Springs, FL 32701) appears to have one of the lowest cost satellite terminals available for the consumer. At \$5000, not including the parabolic dish antenna, it still seems expensive. When compared to similar systems available from competitors, it is downright cheap! Their comprehensive descriptive brochure is very informative.

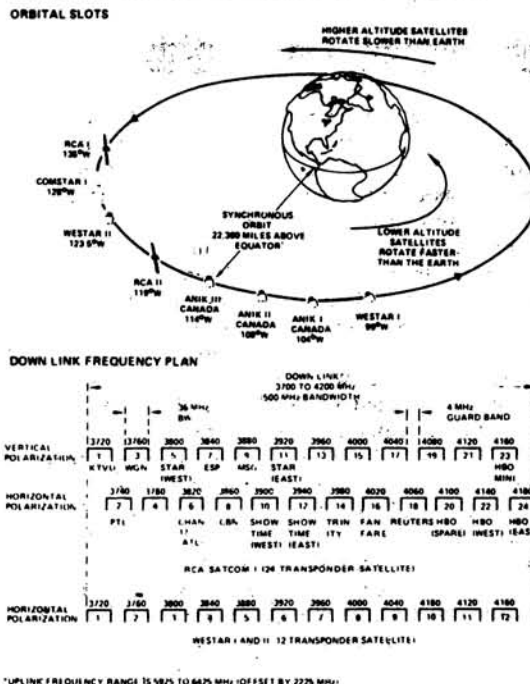
Hood's Lightworks (P.O. Box 1026, Beaverton, OR 97005) provides a complete satellite antenna-pointing chart custom-printed for your geographical location for only \$4. Our sample provided 20 separate listings for present and prospective satellites.

Catalogs and literature from manufacturing giants like *Scientific Atlanta* (3845 Pleasantdale Road, Atlanta, GA 30340) and *American Satellite Corporation* (20301 Century Blvd., Germantown, MD 20767) provide technical insights into manufactured products designed for professional applications.

Some readers have undoubtedly heard of Japanese home TV terminals in the \$500 range. It's true. But both the Japanese technology and geographical demands are much different from ours. Japanese TV satellites operate in the 12 gigaHertz (12,000 MHz) range, making signal-beaming sharper. Additionally, the Japanese mainland is hardly larger than the states of Florida and Georgia. Thus, while our satellite power is spread over much of the North American continent, the Japanese satellite may beam all of its energy on one concentrated spot, making signal levels there much higher. Under those circumstances, receiving terminals become much less demanding. Smaller antennas and less critical amplifiers are easily constructed at lower prices.

At the present time, a number of manufacturers are offering "turnkey"

This chart of satellite positions and frequencies (courtesy of Hughes Microwave Communications Products) shows the up-link (earth-to-satellite) frequencies and those for the down-link (satellite-to-earth). All of the satellites used for commercial reception are in "stationary" (relative to the earth's rotation) orbits so continual tracking and antenna adjustment is not necessary.



satellite receiving terminals. They will set up the equipment at your location, and you take it from there. New systems manufacturers are developing all the time, and prices will come down competitively. As more equipment becomes available at more attractive prices, consumer demand will increase. This will result in increased production driving prices down further. Clearly, home satellite TV reception is upon us.

A Fly in the Ointment. There is still one discord in planning your own home TV terminal: FCC licensing. Yes, even though your prospective installation is for *receiving only*, and will never transmit, and even though that TV reception is for your own entertainment, not to be distributed with or without profit to

others, you must be licensed.

Years ago, the Federal Communications Commission ruled that all microwave links must be licensed, both at the transmitting and receiving ends. This was done in order to register the locations of all antennas to avoid future transmitting installations which might interfere with prior licensees.

Needless to say, not all home TV users are licensed, but the law is specific—at least for now. Pressure is being brought to bear on the FCC to deregulate the satellite TV receive-only terminals, and many in the industry say it is only a matter of time. Even now, the licensing procedure is quite routine.

Today, most of us are watching a limited number of local channels, but tomorrow...

Directional Antennas

□ Most of us approach the job of shortwave antenna installation in a casual manner; a long piece of wire is run away from obstructions and power lines and as high as possible. Hams are frequently guilty of the same sort of cavalier attitude toward antennas: "If it works, leave it up; if it doesn't, put

up something else."

The fact of the matter is that shortwave receiving antennas, like other kinds of antennas, must be properly designed for optimum performance, and the job is not all that difficult once we understand a few of the underlying principles. Fortunately, shortwave re-

ceiving antennas are far more forgiving of sloppy design than are their transmitting counterparts. They don't balk at high SWR, they don't require special insulation considerations, and they needn't be concerned with large diameter conductors to accommodate radio frequency power. A receiving an-

tenna is merely a conductor in the path of an arriving field of weak electromagnetic energy, and nothing more.

Back To Basics. In order to intercept the RF energy efficiently, a number of considerations are necessary. Like transmitting antennas, some lengths of antenna are more suited to certain radio frequencies than others. That is to say, certain lengths "resonate," or reinforce, radio signals whose wavelengths are mathematically related to the electrical length of the antenna wire or conductor. This is why CB antennas are all carefully adjusted to optimize their performance in the 27 MHz band.

Unlike CB, shortwave reception is very much broadband, typically from 3 to 30 MHz. This ten-to-one frequency range means that a resonant antenna is out of the question. This is where antenna "tuners" or "matchboxes" often enter the picture, especially for transmitting. From a practical standpoint, however, tuning devices are rarely necessary for receiving installations. A properly designed receiving antenna will work quite well over a wide frequency range, as evidenced by the reasonably good performance of that random length of wire discussed earlier.

Omnidirectionals. For the casual shortwave listener, desirable signals arrive from all points of the compass. He or she is unconcerned with the prospect that two signals arriving simultaneously on the same frequency make listening rather difficult. If one can't hear the station, it's merely a move up or down the dial for a new quarry.

In most cases, a random wire antenna, no more than 20 to 30 feet in length, will provide excellent worldwide reception. As with any antenna design, it should be as high as possible, and far away from physical obstructions—especially those which are metallic—and clear of power lines. Such random antennas may even be run at angles, following the contours of a roof line or building eave.

As an alternate solution, a very useful receiving antenna may be made from twenty feet of inexpensive mast pipe or conduit. It is mounted vertically, right on a ground pipe, and insulated from it by a sleeve of PVC pipe. A length of coax cable (any standard type) is run from the antenna to the receiver. Single vertical elements are always omnidirectional. Such an antenna configuration will give an excellent accounting of itself throughout the shortwave spectrum.

Directionals. A trip to an FCC monitoring station or to an intelligence

gathering electronic surveillance installation will awe any but the most jaded listener. The Wullenweber antenna arrays they use look like elephant compounds. Is it really necessary to take such elaborate measures to improve shortwave reception? Certainly not. These electronic behemoths are designed with a very special purpose in mind: to measure exact bearings on incoming signals over wide ranges of frequencies.

However, we can learn a lesson from these sophisticated installations. With proper design, an antenna can be made directional, and some of these designs may be simpler than you ever dreamed.

The Dipole. The simplest of all directional antennas is the dipole. Usually center-fed by a length of coaxial cable, the radiation (and reception) pattern of a dipole is that of a dough-

nut. Major receiving strength extends broadside to the antenna, which is relatively insensitive to signals arriving off the ends. In fact, deep nulls reject signals exactly off the ends of the dipole, making it very useful for interference cancellation from power lines or local transmitters. Since the reception lobes are very broad and the nulls are very sharp, it is better to point the axis of the antenna at the source of interference in order to reduce it, instead of worrying about the broadside of the antenna facing a particular geographical location to favor reception.

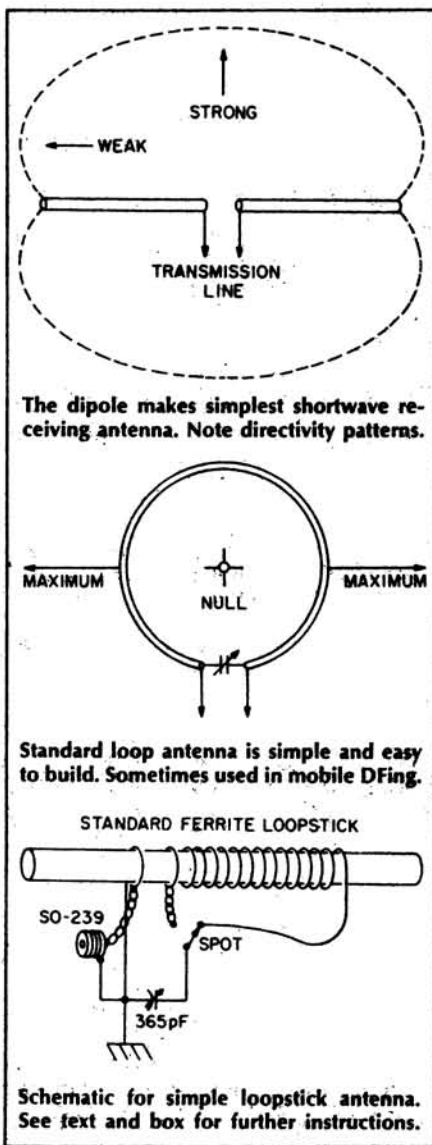
Arriving radio signals are rarely stable targets. Only when the transmitter is relatively close to the receiver can the listener be assured of constant directional reception. The reason for this is the characteristic differences between ground waves and sky waves. Ground waves tend to travel in a straight path close to the surface of the earth, while sky waves are warped by the ionosphere, an electrically-charged layer of our atmosphere hundreds of miles above us. When these waves return to Earth, they are grossly altered from their original symmetrical pattern and arrive at various angles.

A true directional antenna, such as that used for radio direction finding purposes and radio navigation, is carefully constructed to reject sky waves and concentrate on ground waves. Still, we can suppose generally that since the waves all started from one location, we can expect that *most* of the arriving signal will come from that direction, broad as it may be.

Beams. Commonly called the "beam" antenna, the Yagi-Uda array is the most common directional antenna encountered in ham radio installations, and it works well. Unfortunately, the Yagi has a very narrow bandwidth of operation, and the lower the frequency, the larger the array necessary. For this reason, Yagis are rarely used solely countered below 14 MHz in hobby radio installations.

Quads. Large and ungainly, the cubical quad is popular among hams and CBers. The basic rules of Yagi element spacing apply to the quad. Like the Yagi, it is not practical as a shortwave receiving-only antenna, due to its size, complexity and fragility. It does, however, offer excellent directivity and gain factors, especially in two-way usage.

Log Periodics. The log periodic dipole array is used extensively for worldwide communications.



While useful over a 3:1 or even 4:1 frequency range, the elements of these broadband antennas are extremely large at the lower shortwave frequencies. Some of these antennas are hundreds of feet in length at major shortwave communications installations! Obviously, the most common and practical use of the LPDA is on VHF and UHF.

But the picture is not as bleak as it appears. There are some tricks the SWL can use which provide excellent antenna directivity, while utilizing small size and inexpensive construction methods and materials.

available with plug-in ferrite loops to cover 10 kHz through 5.5 MHz. Because the loop is tiltable, skywave anomalies may be greatly reduced, improving the directivity of the device.

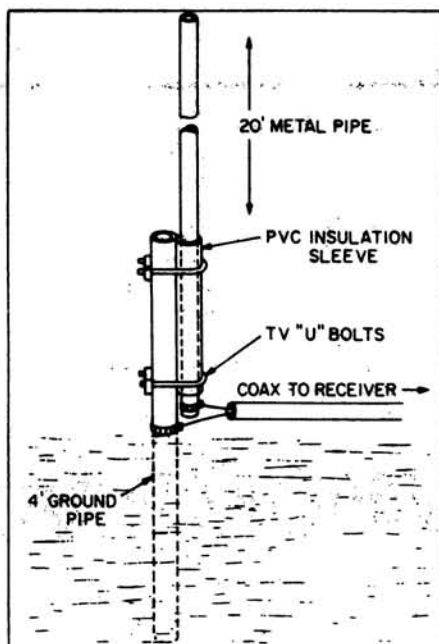
The Palomar loop antenna also has a built-in amplifier. This is important for two reasons: the short antenna doesn't capture much signal, and the impedance of a loop is very low, making a poor match when connected directly to a receiver.

A Homebrew Loop. For the home experimenter, a ferrite rod loop antenna may be easily constructed once the materials are located. The experimenter will need a standard ferrite rod about seven-inches long, two rubber grommets to fit over the rod, a small aluminum utility box, an SPDT switch, a miniature 365-pF variable capacitor, a few feet of insulated hookup wire, and an SO-239 female coax connector.

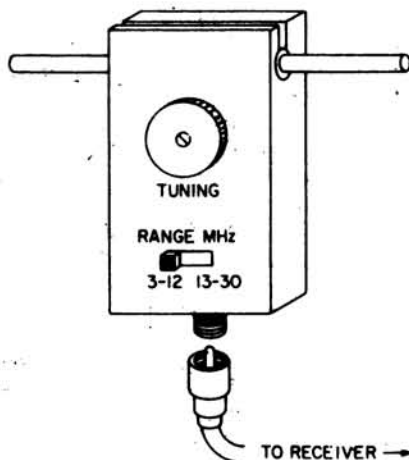
The ferrite rod is wound tightly at its center with a single layer of cotton-covered hookup wire. Start by leaving about four-inches of wire as a beginning pigtail (this will be the ground lead). After one turn, twist tightly together another four inches of wire; this will be the output link. Resume winding another single turn and twist tightly together another four inches of wire; this lead will go to one side of the frequency range switch. Resume winding again until another 13 turns have been completed bringing the total to 15

The Loop. Probably the most common directional low frequency antenna is the loop. Originally developed for direction finding service in the maritime industry, a loop antenna has a figure 8 pattern. Its null is at right angles to the plane of the loop, and its maximum signal pickup is off the ends.

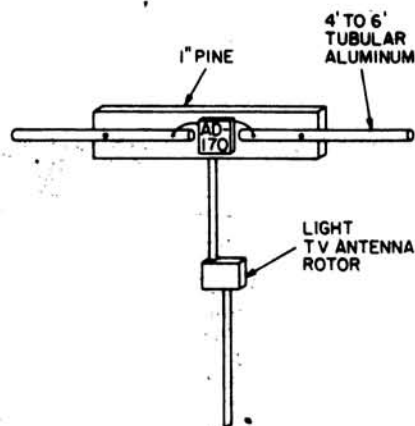
A more recent variation on the receiving loop is the ferrite rod loop. One commercial unit is manufactured by



A good ground connection is a must for shortwave reception. Helps in storms, too.



This is the way a completed loopstick antenna should look. It's a good little unit.



The Datong Active Antenna, Model AD-170 fits within dipole antenna elements!

Palomar Engineers. The Palomar unit is turns. Leave about four-inches of pigtail to be connected to the other side of the frequency range switch.

Tape the coil down firmly against the rod to hold it in place. Slip a rubber grommet over each end of the rod to act as a cushion when it is fitted in place in the aluminum box.

Drill two holes through opposite sides of the utility box to accommodate the rubber grommets. With a hacksaw, cut a slot the width of the grommet holes from the top of the box all the way down to the holes. This will provide a Faraday shield to reduce the capacitive effects of nearby objects on the pattern of the loop, as well as serve as a mounting area for the ferrite rod. Be sure to leave room for the switch, jack, and tuning capacitor.

Cut and drill additional holes to allow for mounting of the variable tuning capacitor, SO-239 chassis connector, range switch, and any mounting bracket for a handle or brace.

In actual operation, the range switch selects the tuning range of the antenna, and the 365-pF tuning capacitor peaks the incoming signal. When properly adjusted, the ferrite loop directional antenna will exhibit a very sharp null when pointed like a gun barrel at the source of the signal. It makes a good direction finding antenna for locating hidden transmitters anywhere in the shortwave spectrum.

The same hazards associated with unpredictable readings from skywaves, as mentioned earlier, still apply. Readings are best when sources are close, and when operating frequencies are low. Tilting the antenna up or down will frequently enhance the sharp directivity of the loop and increase accuracy on bearings. This is the same technique used with the Palomar unit.

Active Antennas. In an effort to overcome the lack of appeal of large antennas for maximum signal pickup, the idea of a built-in preamplifier sounds attractive. In fact, several manufacturers now offer short receiving antennas which detect only a small amount of energy and amplify it, sending a husky signal to the receiver. Often, these antennas may be mounted in such a way as to null out electrical interference, favoring the desirable signals.

Antenna Rotation. Most antennas can be rotated, even those monsters used by the military services. But most of us have a balanced budget, and are con-

cerned about deficit spending. For those reasons, as well as the possibility of partially rotating the house under the antenna, we shall confine our discussion to small directional antennas.

The trouble with the remote positioning of the ferrite loop is its inaccessibility for tuning.

A better approach is the use of a broadband antenna. Experimentally, we tried a waterproofed version of Da-

tong's popular AD-170. It is capable of continuous 70 kHz to 70 MHz performance. It comes from the factory with wire leads for its dipole elements.

By mounting the preamplifier against a sturdy wood or plastic brace, and replacing the wire elements with two lengths of self-supporting aluminum tubing at least four-feet long per leg, the assembly makes a good rotatable directional shortwave antenna. The problem

of skywaves distorting arriving waves will still make true directivity a little difficult to obtain.

Conclusion. While any hunk of wire is theoretically capable of intercepting radio energy, a little planning can make a big difference. By analyzing the requirements of your particular receiving installation, an antenna can be designed and installed with expectations of excellent performance. ■

Shortwave Broadcasts

□ At midnight GMT on Tuesday, January 23, 1973, the World Service of the British Broadcasting Corporation announced that Lyndon Baines Johnson, 36th President of the United States, had passed away at his home near Johnson City, Texas. On Friday, August 9, 1974, at 1500 hours GMT, station CFRX in Toronto, Ontario, Canada, stated that within one hour Gerald R. Ford would become President of the United States.

On Wednesday, March 26, 1975, at 0100 GMT, HCJB in Quito, Ecuador, reported that King Faisal of Saudi Arabia, while holding court in Riyadh, had been assassinated by his nephew, Prince Musad. Where did I get this information? Did I write to broadcasters and ask when and how they had reported various historical events? Did I travel to England and take notes while peering over the shoulders of the professional SWL's at the BBC's Caversham Park monitoring station?

The answer to both these questions is "no." To obtain the facts that open this article, I merely looked in my log books, located in my den, in my house. Keeping track of the opening stories of shortwave news programs has become a "hobby-within-a-hobby" for me.

Log Sheets. Compiling a "news-log" is an easy task. Using commercially prepared logging pages, such as those provided by Gilfer Shortwave, in Park Ridge, NJ, the listener has a few inches of space in every line to make notes on. This is in addition to the usual time, date, and frequency columns. The Gilfer forms are pre-punched to fit in a three-hole, loose-leaf notebook. Other companies also market logging materials, or they can easily be made at home using ruler and pen or pencil.

Simple receiving equipment is frequently enough to pursue this "spinoff" hobby. Radio Canada International

Radio Nederland, Deutsche Welle, HCJB, and other stations that send strong signals into North America, begin their transmissions with newscasts. These broadcasts can usually be heard on inexpensive portable receivers.

If you don't make your letters too large, you'll have space to jot a few brief notes about every newscast you hear. Recent notations in my log include such things as: "200 dead in Yugoslav earthquake;" "Whereabouts of Amin still a mystery;" and "U.S. ends economic aid to Pakistan over nuclear development." Each notation fits nicely into the four-and-three-quarter-inch "Program/Contact" space on a Gilfer LS-1 logging sheet.

The listener will soon have his or her own personal chronicle of war, disaster, pestilence, revolution—all the occurrences that comprise the news of the day. It's often fascinating to examine the stories various stations use to open their newscasts and compare them with items carried by other broadcasters. Also, hearing one story as reported by several sources can often contribute to the listener's understanding of it.

The record will include more than world events. Hearing what a foreign

shortwave broadcaster says about happenings in your own country is frequently a good way to get information.

Following another country's "local" stories from day to day can be intriguing, too. Election campaigns are examples of this type of story, as are reports of natural disasters. How Darwin, Australia, recovered from the cyclone that devastated it on Christmas Day, 1974 was captivating news.

Student SWLs preparing papers can have knowledge of precisely when recent events occurred. All they need do is consult their own listening post logs.

Stories that develop over a long period of time, such as the events leading to the resignation of Richard Nixon, can be examined later to find out what happened in a daily parade of historical occurrences.

By writing just a few words during each news program he or she hears, the SWL will soon have a readily available volume of recent world history. And, the longer the listener keeps such records, the greater the time period covered by this individualized history book. ■

News log forms like this one are available from Gilfer Shortwave, or you can make them up on your own. Keeping them in some sort of binder gives you an ongoing collection of world events that can be reviewed and updated as events develop & take shape.

DATE	TIME	FREQUENCY	STATION	PROGRAM	REMARKS	NEW	NEW	NEW
MM/DD	HH:MM	KHz	Call	Time	Comments	STATION	STATION	STATION
1/23/73	11:00	11.715	BB	11:00	News program from London			
1/23/73	11:05	11.715	BB	11:05	Continuation of news program			
1/23/73	11:10	11.715	BB	11:10	Continuation of news program			
1/23/73	11:15	11.715	BB	11:15	Continuation of news program			
1/23/73	11:20	11.715	BB	11:20	Continuation of news program			
1/23/73	11:25	11.715	BB	11:25	Continuation of news program			
1/23/73	11:30	11.715	BB	11:30	Continuation of news program			
1/23/73	11:35	11.715	BB	11:35	Continuation of news program			
1/23/73	11:40	11.715	BB	11:40	Continuation of news program			
1/23/73	11:45	11.715	BB	11:45	Continuation of news program			
1/23/73	11:50	11.715	BB	11:50	Continuation of news program			
1/23/73	11:55	11.715	BB	11:55	Continuation of news program			
1/23/73	12:00	11.715	BB	12:00	Continuation of news program			
1/23/73	12:05	11.715	BB	12:05	Continuation of news program			
1/23/73	12:10	11.715	BB	12:10	Continuation of news program			
1/23/73	12:15	11.715	BB	12:15	Continuation of news program			
1/23/73	12:20	11.715	BB	12:20	Continuation of news program			
1/23/73	12:25	11.715	BB	12:25	Continuation of news program			
1/23/73	12:30	11.715	BB	12:30	Continuation of news program			
1/23/73	12:35	11.715	BB	12:35	Continuation of news program			
1/23/73	12:40	11.715	BB	12:40	Continuation of news program			
1/23/73	12:45	11.715	BB	12:45	Continuation of news program			
1/23/73	12:50	11.715	BB	12:50	Continuation of news program			
1/23/73	12:55	11.715	BB	12:55	Continuation of news program			
1/23/73	13:00	11.715	BB	13:00	Continuation of news program			
1/23/73	13:05	11.715	BB	13:05	Continuation of news program			
1/23/73	13:10	11.715	BB	13:10	Continuation of news program			
1/23/73	13:15	11.715	BB	13:15	Continuation of news program			
1/23/73	13:20	11.715	BB	13:20	Continuation of news program			
1/23/73	13:25	11.715	BB	13:25	Continuation of news program			
1/23/73	13:30	11.715	BB	13:30	Continuation of news program			
1/23/73	13:35	11.715	BB	13:35	Continuation of news program			
1/23/73	13:40	11.715	BB	13:40	Continuation of news program			
1/23/73	13:45	11.715	BB	13:45	Continuation of news program			
1/23/73	13:50	11.715	BB	13:50	Continuation of news program			
1/23/73	13:55	11.715	BB	13:55	Continuation of news program			
1/23/73	14:00	11.715	BB	14:00	Continuation of news program			
1/23/73	14:05	11.715	BB	14:05	Continuation of news program			
1/23/73	14:10	11.715	BB	14:10	Continuation of news program			
1/23/73	14:15	11.715	BB	14:15	Continuation of news program			
1/23/73	14:20	11.715	BB	14:20	Continuation of news program			
1/23/73	14:25	11.715	BB	14:25	Continuation of news program			
1/23/73	14:30	11.715	BB	14:30	Continuation of news program			
1/23/73	14:35	11.715	BB	14:35	Continuation of news program			
1/23/73	14:40	11.715	BB	14:40	Continuation of news program			
1/23/73	14:45	11.715	BB	14:45	Continuation of news program			
1/23/73	14:50	11.715	BB	14:50	Continuation of news program			
1/23/73	14:55	11.715	BB	14:55	Continuation of news program			
1/23/73	15:00	11.715	BB	15:00	Continuation of news program			
1/23/73	15:05	11.715	BB	15:05	Continuation of news program			
1/23/73	15:10	11.715	BB	15:10	Continuation of news program			
1/23/73	15:15	11.715	BB	15:15	Continuation of news program			
1/23/73	15:20	11.715	BB	15:20	Continuation of news program			
1/23/73	15:25	11.715	BB	15:25	Continuation of news program			
1/23/73	15:30	11.715	BB	15:30	Continuation of news program			
1/23/73	15:35	11.715	BB	15:35	Continuation of news program			
1/23/73	15:40	11.715	BB	15:40	Continuation of news program			
1/23/73	15:45	11.715	BB	15:45	Continuation of news program			
1/23/73	15:50	11.715	BB	15:50	Continuation of news program			
1/23/73	15:55	11.715	BB	15:55	Continuation of news program			
1/23/73	16:00	11.715	BB	16:00	Continuation of news program			
1/23/73	16:05	11.715	BB	16:05	Continuation of news program			
1/23/73	16:10	11.715	BB	16:10	Continuation of news program			
1/23/73	16:15	11.715	BB	16:15	Continuation of news program			
1/23/73	16:20	11.715	BB	16:20	Continuation of news program			
1/23/73	16:25	11.715	BB	16:25	Continuation of news program			
1/23/73	16:30	11.715	BB	16:30	Continuation of news program			
1/23/73	16:35	11.715	BB	16:35	Continuation of news program			
1/23/73	16:40	11.715	BB	16:40	Continuation of news program			
1/23/73	16:45	11.715	BB	16:45	Continuation of news program			
1/23/73	16:50	11.715	BB	16:50	Continuation of news program			
1/23/73	16:55	11.715	BB	16:55	Continuation of news program			
1/23/73	17:00	11.715	BB	17:00	Continuation of news program			
1/23/73	17:05	11.715	BB	17:05	Continuation of news program			
1/23/73	17:10	11.715	BB	17:10	Continuation of news program			
1/23/73	17:15	11.715	BB	17:15	Continuation of news program			
1/23/73	17:20	11.715	BB	17:20	Continuation of news program			
1/23/73	17:25	11.715	BB	17:25	Continuation of news program			
1/23/73	17:30	11.715	BB	17:30	Continuation of news program			
1/23/73	17:35	11.715	BB	17:35	Continuation of news program			
1/23/73	17:40	11.715	BB	17:40	Continuation of news program			
1/23/73	17:45	11.715	BB	17:45	Continuation of news program			
1/23/73	17:50	11.715	BB	17:50	Continuation of news program			
1/23/73	17:55	11.715	BB	17:55	Continuation of news program			
1/23/73	18:00	11.715	BB	18:00	Continuation of news program			
1/23/73	18:05	11.715	BB	18:05	Continuation of news program			
1/23/73	18:10	11.715	BB	18:10	Continuation of news program			
1/23/73	18:15	11.715	BB	18:15	Continuation of news program			
1/23/73	18:20	11.715	BB	18:20	Continuation of news program			
1/23/73	18:25	11.715	BB	18:25	Continuation of news program			
1/23/73	18:30	11.715	BB	18:30	Continuation of news program			
1/23/73	18:35	11.715	BB	18:35	Continuation of news program			
1/23/73	18:40	11.715	BB	18:40	Continuation of news program			
1/23/73	18:45	11.715	BB	18:45	Continuation of news program			
1/23/73	18:50	11.715	BB	18:50	Continuation of news program			
1/23/73	18:55	11.715	BB	18:55	Continuation of news program			
1/23/73	19:00	11.715	BB	19:00	Continuation of news program			
1/23/73	19:05	11.715	BB	19:05	Continuation of news program			
1/23/73	19:10	11.715	BB	19:10	Continuation of news program			
1/23/73	19:15	11.715	BB	19:15	Continuation of news program			
1/23/73	19:20	11.715	BB	19:20	Continuation of news program			
1/23/73	19:25	11.715	BB	19:25	Continuation of news program			
1/23/73	19:30	11.715	BB	19:30	Continuation of news program			
1/23/73	19:35	11.715	BB	19:35	Continuation of news program			
1/23/73	19:40	11.715	BB	19:40	Continuation of news program			
1/23/73	19:45	11.715	BB	19:45	Continuation of news program			
1/23/73	19:50	11.715	BB	19:50	Continuation of news program			
1/23/73	19:55	11.715	BB	19:55	Continuation of news program			
1/23/73	20:00	11.715	BB	20:00	Continuation of news program			
1/23/73	20:05	11.715	BB	20:05	Continuation of news program			
1/23/73	20:10	11.715	BB	20:10	Continuation of news program			
1/23/73	20:15	11.715	BB	20:15	Continuation of news program			
1/23/73	20:20	11.715	BB	20:20	Continuation of news program			
1/23/73	20:25	11.715	BB	20:25	Continuation of news program			
1/23/73	20:30	11.715	BB	20:30	Continuation of news program			
1/23/73	20:35	11.715	BB	20:35	Continuation of news program			
1/23/73	20:40	11.715	BB	20:40	Continuation of news program			
1/								

Understanding Electronics

Amps, Ohms and Volts

□ ONE OF THE MOST thought-provoking discoveries of modern physics is the fact that matter and energy are interchangeable. Centuries of scientific head-scratching about the nature of matter, the mystery of fire, and the once-terrifying crack of lightning have all come to focus on the smallest particle that is the building block of any given substance: the atom. An atom is necessarily matter and yet this atom of matter can undergo nuclear fission and release quantities of energy that are beyond the imagination. In the atom lies the secret of all phenomena. One theory of the universe, hypothesized by Georges Lemaitre, even regards the present universe as resulting from the radioactive disintegration of one primeval atom!

By the beginning of the 19th century, the atomic theory of matter—which actually originated in 5th century Greece when the atom was named—was firmly established. It was due primarily to the efforts of 17th century scientists who—actually working in the tradition of medieval alchemy—sought the prime constituent of all matter. Mainly through the work of John Dalton, whose investigations as to how various elements combine to form chemical compounds, it came to be regarded that an atom was the indivisible and indestructible unit of matter.

This viable and working view of the indestructible atom served science until 1897 when the atom itself was found to be destructible! To anyone concerned with electricity or electronics, the year 1897 is a memorable one: it was the year J. J. Thomson, the English physicist, identified and experimentally revealed the existence of the first subatomic particle—the *electron*!

The First "Electronic" Experiment. We blithely speak of electricity as the flow of electrons yet, often, we are little aware of the great body of research that went into elucidating this fundamental of basic electricity. In fact, before the discovery of the electron, convention held that the flow of electric current was in the direction that a positive charge moved. This convention of *positive current*, being the flow of positive charges

and opposite to the direction of electron flow, is still found to be useful in circuit analysis and is used even today.

Thomson's experiment established that a particle much lighter than the lightest atom did indeed exist. The electron, as it was named, was the first subatomic particle to be defined.

The experiment was conducted utilizing a rudimentary version of a cathode ray tube—the modern version of which is in almost every home today in the form of the television picture tube. Before Thomson's experiment, it was discovered that when electric current was passed through a gas in a discharge tube, a beam of unknown nature traveled through the tube from the negative to positive terminal (opposite to the direction conventionally held as the direction of the flow of current).

This "cathode ray" beam also traveled in a straight line and was deflected by electric or magnetic forces applied perpendicular to the beam. What Thomson did was to use these facts to determine for one of the mysterious particles comprising the beam of cathode rays the relationship of its mass, m , to its electric charge, e . By deflecting the beam with a known electric force (Fig. 1) and then measuring what magnetic force applied in the opposite direction would bring the beam back to its original undeflected position, he could determine the relationship of e to m . He established a definite value for e/m and thereby "discovered" the electron which, as we now know, is 1,837 times smaller in mass than the lightest atom, the hydrogen atom. It also carries the smallest charge that occurs in nature; every electric charge is actually an integral multiple of the charge of the electron.

From Minus to Plus. With the discovery of the electron, it was still over a dozen years into the 20th century before a graphic conception of the atom evolved. Since the atom is electrically neutral and electrons are negatively charged, the existence of positively charged particles was a necessity, and the existence of a *proton* was postulated. Eventually the nuclear model of the atom was evolved. Each atom was

conceived to resemble a solar system in miniature. The nucleus—positively charged—is surrounded by a number of electrons revolving around it; the charges balance and the atom is electrically neutral (Fig. 2). Further research in the 20th century has gone on to reveal more elementary particles than you can shake a stick at: neutrons, positrons, neutrinos, mesons, and more. The number continues to grow and yet the ultimate nature of matter remains a riddle. But, in a discussion of basic electricity, only the electron and proton need concern us.

Electrons in Orbit. An atom of matter has a number of electrons orbiting around its nucleus. A hydrogen atom,

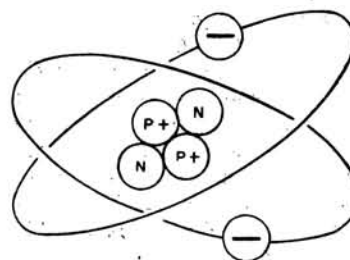


Fig. 2. Charge of each electron balances that of a proton. Other particles affect atomic mass but can be ignored in study of electronics.

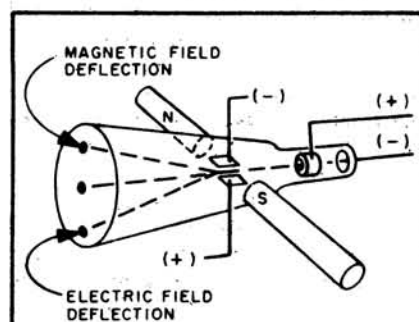


Fig. 1. Electron beam, like that in a TV picture tube (CRT), can be deflected magnetically or by an electric field. Force needed "measured" the electron.

Understanding Electronics

for example, has a single electron; carbon on the other hand has 6. These electrons are arranged in rings or shells around the central nucleus—each ring having a definite maximum capacity of electrons which it can retain. For example, in the copper atom shown in Fig. 3 the maximum number of electrons that can exist in the first ring (the ring nearest the nucleus) is two. The next ring can have a maximum of eight, the third ring a maximum of 18, and the fourth ring a maximum of 32. However, the outer ring or shell of electrons for any atom cannot exceed eight electrons. However, heavier atoms may have more than four rings.

The Outer Orbit. The ring of electrons furthest from the atom's nucleus is known as the *valence ring* and the electrons orbiting in this rings are known as *valence electrons*. These valence electrons, being further from the nucleus, are not held as tightly in their orbits as electrons in the inner rings and can therefore be fairly easily dislodged by an external force such as heat, light, friction, and electrical potential. The fewer electrons in the valence ring of an atom, the less these electrons are bound to the central nucleus. As an example, the copper atom has only one electron in its valence ring. Consequently, it can be easily removed by the application of only the slightest amount of external energy. Ordinary room temperature is sufficient to dislodge large numbers of electrons from copper atoms; these electrons circulate about as free electrons. It is because of these large numbers of free electrons that copper is such a good electrical conductor. There could be no electrical or electronics industry as we know it today if it were not for the fact that electrons can fairly easily escape, or be stripped from the valence ring of certain elements.

Electronic Charges. If an electron is stripped from an atom, the atom will assume a positive charge because the number of positively charged protons in its nucleus now exceeds the number of negatively charged orbiting electrons. If, on the other hand, the atom should gain an electron, it will become negatively charged as the number of electrons now exceeds the protons in its nucleus. The atom with the deficiency of electrons is known as a *positive ion*, while an atom with a surplus of electrons is known as a *negative ion*.

Presence of an electrical charge on a body can be illustrated by use of an electroscope (Fig. 4). Two leaves of aluminum or gold foil hang from a

metal rod inside a glass case so they're free from air disturbances. When the metal rod is touched by a charged body, the leaves acquire static electricity of the same polarity and, since like charges repel, they stand apart. The greater the charge, the further apart the leaves spread.

Electron Flow. When an electrical conductor is placed between these two oppositely charged bodies, free electrons are attracted by the positive body—free electrons will move through the wire. This movement of free electrons will continue only until the excess of electrons is equally divided between the two bodies. Under these conditions, the charges on both bodies will be equal and the electron flow will end.

In Fig. 5 are a battery, lamp, and connecting leads between the battery and lamp. In this instance, the battery serves as an electric charge pump—free electrons continually developed at its negative terminal by chemical action flow through the connecting leads and lamp back to the positive terminal of the battery by the attraction of oppositely charged bodies. The battery, connecting leads, and lamp form an electrical circuit which must be complete before the free electrons can flow from the battery's negative terminal to its positive terminal via the lamp. Thus, the battery serves as a source of potential difference or voltage by continually supplying a surplus of electrons at its negative terminal. Summing up, we can say a flow of electric current consists of the movement of electrons between two oppositely charged bodies.

We cannot progress very far into the study of electricity without first becoming familiar with the basic properties of electrical circuits. Just as we define dis-

tance in feet and inches, so do we define electrical properties in specific terms and units.

Potential. Earlier, we saw that an electric charge difference has to exist between the ends of an electrical conductor in order to cause a flow of free electrons through the conductor. This flow of electrons constitutes the electric current. The electric charge difference, or potential difference, exerts a force on the flow of free electrons, forcing them through the conductor. This electric force or pressure is referred to as *electromotive force*, abbreviated *EMF*.

The greater the charge or potential difference, the greater will be the movement of free electrons (current) through the conductor as there will be more "push and pull" on the free electrons. The symbol used to designate electrical potential is the letter *E* which stands for *electromotive force*. The quantity of EMF is measured by a unit called the volt. Hence, the common name most often used in place of EMF is *voltage*.

Current Intensity. We have learned that an electric current consists of a flow of charge carriers (generally free electrons) between two points of different electrical potential. The rate of flow

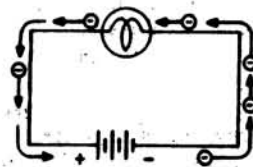


Fig. 5. Electron flow in any circuit is from negative to positive.

Fig. 3. The number of electrons to each ring are limited—2 in first; 8 in second; 18 for the third; and a total of 32 in fourth orbital ring.

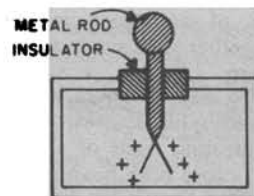
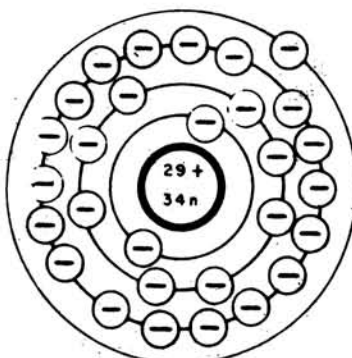


Fig. 4. Electroscope is a simple device to indicate electrical charges that are too weak to be measured with standard meters.

of these changes determines the intensity or strength of this current flow. Current strength is expressed in units known as *amperes*. One ampere of current flows in a circuit when 6,240,000,000,000,000 electrons flow out of a negative terminal, through a conductor, and back into a positive terminal in one second. The symbol for the ampere is the letter *I* which stands for intensity.

Resistance. The flow of electric current through a conductor is caused by the movement of free electrons present in the atoms of the conductor. A bit of thought then indicates that the greater the number of free electrons present in the atoms of a particular conductor, the greater will be its electrical conductivity. Gold, silver, and copper rank as excellent electrical conductors, as their atoms readily release free electrons. On the other hand, the atoms of such elements as sulphur have almost no free electrons available and they are thus very poor electrical conductors. Such materials are known as electrical insulators. Between these extremes lie elements such as carbon whose atoms have a moderate number of free electrons available and thus are moderately good electrical conductors.

Even the best electrical conductors offer some opposition to the passage of free electrons. This opposition is called resistance. You might consider electrical resistance similar to mechanical friction. As in the case of mechanical friction, electrical resistance generates heat. When current flows through a resistance, heat is generated; the greater the current flow, the greater the heat. Also, for a given current flow, the greater the resistance, the greater the heat produced.

Electrical resistance can be both beneficial and undesirable. Toasters, electric irons, etc. all make use of the heat generated by current flowing through wire coils. Resistance is also often intentionally added to an electrical circuit to limit the flow of current. This type of resistance is generally lumped together in a single unit known as a resistor.

There are also instances where resistance is undesirable. Excessive resistance in the connecting leads of an electrical circuit can cause both heating and electrical loss. The heating, if sufficient, can cause a fire hazard, particularly in house wiring, and the circuit losses are a waste of electrical power.

Electrical resistance is expressed by a unit known as the *ohm*, indicated by the letter *R*. An electrical conductor has a resistance of one ohm when an applied EMF of one volt causes a current of one ampere to flow through it.

Resistance Factors. There are other factors beside the composition of the material that determine its resistance. For example, temperature has an effect on the resistance of a conductor. As the temperature of copper increases, for example, its resistance increases. The increase in temperature causes the electrons in the outer ring of the atom to resist release to the free electron state. This increase in resistance is known as a *positive temperature coefficient*. Not all conductors show this increase in resistance with an increase in temperature; their resistance decreases with an increase in temperature. Such materials are said to have a *negative temperature coefficient*. Certain metallic alloys have been developed which exhibit a *zero temperature coefficient*: their resistance does not change with changes in temperature.

As you might suspect, the length of a conductor has an effect upon its resistance. Doubling the length of a conductor will double its resistance. By the same token, halving the length of a conductor will cut its resistance in half. Just remember that the resistance of a conductor is *directly proportional to its length*.

The cross-sectional area of a conduc-

tor also determines its resistance. As you double the cross-section of a conductor, you halve its resistance; halving its cross-section doubles its resistance. Here again, the "why" of this is pretty easy to see: there are more current-carrying electrons available in a large cross-section conductor than in a small cross-section conductor of the same length. Therefore, the resistance of a conductor is *inversely proportional to its cross-sectional area*.

Circuit Relationship. Now that we have a basic understanding of voltage, current, and resistance, let's take a look at just how they interact under circuit conditions.

Fig. 6A shows a battery, ammeter (a device to indicate current strength), and resistor connected in series. Notice that the ammeter indicates that 4 amperes are flowing in the circuit.

Fig. 6B shows the identical setup with the exception that the battery voltage has now been doubled. The ammeter now shows that twice the original current, or 8 amperes, is now flowing in the circuit. Therefore, we can see that doubling the voltage applied to the circuit will double the current flowing in the circuit.

In Fig. 6C the same circuit appears again; this time, however, the battery voltage is one half its original value. The ammeter shows that one half of the original current, or 2 amperes, is now flowing in the circuit. This shows us that halving the voltage applied to the circuit will halve the current flowing through the circuit.

All this boils down to the fact that, assuming the same circuit resistance in all cases, *the current flowing in a circuit will be directly proportional to the applied voltage*—increasing as the voltage is increased, and decreasing as the applied voltage is decreased.

In Fig. 7A we again see the circuit consisting of the battery, ammeter, and resistance. Notice that the ammeter indicates that 4 amperes are flowing through the circuit.

In Fig. 7B we see that the value of resistance has been cut in half and as a result, the ammeter indicates that twice the original current, or 8 amperes, is now flowing in the circuit. This leads us to the correct assumption that for a given supply voltage, halving the circuit resistance will double the current flowing in the circuit.

Fig. 7C again shows our basic circuit, but with the resistance now doubled from its original value. The ammeter indicates that the current in the circuit is now one half of its original value.

Fig. 6. In A, B, and C, the value of the resistor remains constant while the supply voltage is altered with a resulting current change.

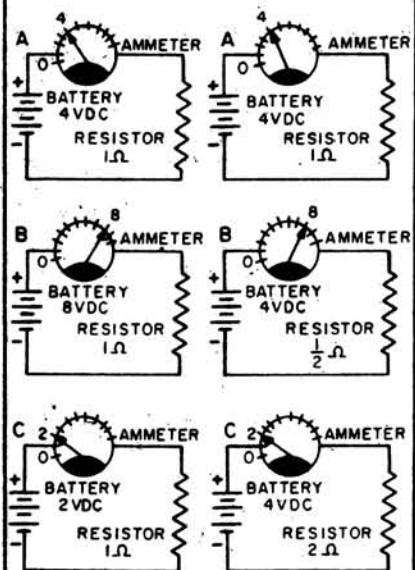


Fig. 7. Battery voltage A, B, and C is held constant while resistor is halved and doubled in value. Resulting current changes are basis for Ohm's law.

Understanding Electronics

Summing things up: for a given supply voltage, the current flowing in a circuit will be inversely proportional to the resistance in the circuit.

Ohm's Law. From what you have seen so far, you are probably getting the idea that you can determine the current flowing in a circuit if you know the voltage and resistance present in the circuit, and the voltage if you know the current and resistance, or the resistance if the voltage and current are known.

All this is quite correct, and is formally stated by Ohm's law as follows:

$$I = \frac{E}{R}$$

Where: E = voltage
I = current
R = resistance

Now, let's take a look at how this formula is used:

To find voltage:

$$E \text{ (voltage)} = I \text{ (current)} \times R \text{ (resistance)}$$

To find current:

$$I \text{ (current)} = \frac{E \text{ (voltage)}}{R \text{ (resistance)}}$$

To find resistance:

$$R \text{ (resistance)} = \frac{E \text{ (voltage)}}{I \text{ (current)}}$$

A handy way to remember Ohm's law is by means of the triangle shown in Fig. 8. Simply cover the quantity (voltage, current, or resistance) that you want to determine, and read the correct relationship of the remaining two quantities. For example, if you want to know the correct current (I),

put your finger over I and read $\frac{E}{R}$.

Covering E or R will yield $I \times R$ or $\frac{E}{I}$ respectively.

Ohm's Law to Determine Voltage.

Let's delve a bit more deeply into Ohm's law by applying it to a few cases where we want to determine the unknown voltage in an electrical circuit. Take a look at Fig. 9, which shows a simple series circuit consisting of a battery and resistor. The value of this resistor is given as 200 ohms, and 0.5 ampere of current is flowing through the circuit. We want to find the value of battery voltage. This is easily done by applying Ohm's

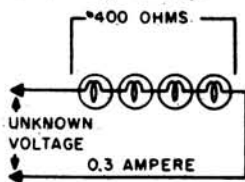


Fig. 10. Although problem looks different the basic circuit is same as that for Fig. 9.

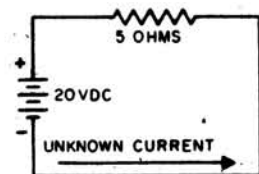


Fig. 12. Basic circuit is same as that in Fig. 11. Although three factors are given, current is unknown quantity.

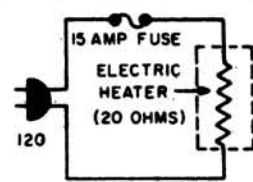


Fig. 11. Formula needed here is different since current is unknown. Just look for triangle in Fig. 8 where "I" is shaded.

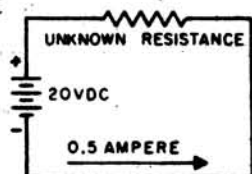


Fig. 13. Most Ohm's law problems are simple series circuits or can be reduced to simple series circuits.

law for voltage as follows:

$$E = I \times R$$

Let's go through this again, this time using a practical illustration. Fig. 10 shows a string of light bulbs, the total resistance of which is 400 ohms. You find that the bulbs draw 0.3 amperes when lighted. Let's say you would like to operate this string of bulbs from the standard 120-volt house current, but you don't know the voltage rating of the individual bulbs. By using Ohm's law for voltage, you can easily determine the voltage to light the bulbs as follows: (unknown voltage) = 0.3 (amperes) x 400 (bulb resistance) = 120 volts.

Ohm's Law to Determine Current.

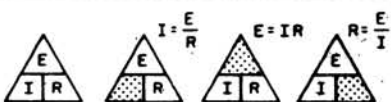


Fig. 8. Shaded portion of triangle indicates unknown quantity in the formula. Visible factors appear in their proper mathematical relation. Just fill in the known values and go on with multiplication or division.

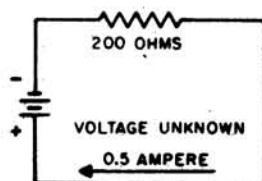


Fig. 9. Unknown quantity, voltage, found easily by applying Ohm's law.

Now, let's take a look at a few examples of how to determine the value of unknown current in a circuit in which both the voltage and resistance are known.

Fig. 11 shows a series circuit with a battery and resistor. The battery voltage is 20 volts DC and the value of resistance is 5 ohms. How much current is flowing through the circuit?

$$\text{Ohm's law for current: } I = \frac{E}{R}$$

$$I \text{ (unknown current)} = \frac{20 \text{ (battery voltage)}}{5 \text{ (resistance in ohms)}}$$

$$I = 4 \text{ amperes}$$

Again to get a bit more practical, let's take a look at Fig. 12. Here we see an electric heater element connected to the 120-volt house line. We know that this particular heater element has a resistance of 20 ohms. The house current line is fused with a 15-ampere fuse. We want to know whether the heater will draw sufficient current to blow the fuse. Here's how to find this out by use of Ohm's law for current.

$$I \text{ (unknown current)} = \frac{120 \text{ (line voltage)}}{20 \text{ (Heater resistance in ohms)}}$$

$$I = 6 \text{ amperes}$$

We find from the above use of Ohm's

law for current that the heater draws 6 amperes, so it can be safely used on the line fused with the 15-ampere fuse. In fact, a 10-ampere fused line can also do the job.

Ohm's Law to Determine Resistance. Ohm's law for resistance enables us to determine the unknown value of resistance in a circuit. Fig. 13 again shows a simple series circuit with the battery voltage given as 20 volts and the current flowing through the circuit as 0.5 ampere. The unknown resistance value in this circuit is found as follows:

$$\text{Ohm's law for resistance: } R = \frac{E}{I}$$

$$R \text{ (unknown resistance)} = \frac{20 \text{ (battery voltage)}}{0.5 \text{ (current in amperes)}}$$

$$R = 40 \text{ ohms}$$

Fig. 14 is a practical example of how to determine unknown resistance. Here, we want to operate a 6-volt light bulb from the 120-volt house line. What value of series dropping resistor do we need to drop the 120-volt house current down to 6 volts? The bulb draws 0.2 ampere.

We must first determine the voltage which must be dropped across the series dropping resistor. This is done by subtracting the line voltage (120) from the bulb's voltage (6). This gives us a value of 114 volts which we use in conjunction with Ohm's law for resistance as follows:

$$R \text{ (unknown resistance)} = \frac{114 \text{ (voltage dropped by resistor)}}{0.2 \text{ (bulb current in amperes)}}$$

$$R = 570 \text{ ohms}$$

Resistance in Series. Many practical electrical and electronic circuits use two or more resistances connected in series.

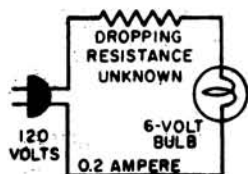


Fig. 14. This Ohm's law problem is somewhat more complex.

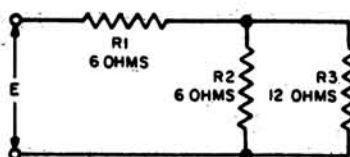


Fig. 17. Ohm's law can be used to determine the equivalent resistance of two or more resistors in parallel. Total current—then solve for ohms.

The point to remember in this case is that the total resistance is the sum of the individual resistances. This is expressed by the formula:

$$R \text{ (total resistance)} = R_1 + R_2 + R_3 + \text{etc.}$$

where R_1 , R_2 , R_3 , etc. are the individual resistances. Thus, in Fig. 15 the total of the individual resistances is $R \text{ (total)} = 40 + 6 + 10 + 5 = 61 \text{ ohms}$.

Resistances may also be connected in parallel in a circuit as in Fig. 16. In this case the current flowing in the circuit will divide between the resistances, the greater current flowing through the lowest resistance. Also, the total resistance in the circuit will always be less than the smallest resistance since the total current is greater than the current in any of the individual resistors. The formula for determining the combined resistance of the two resistors is:

$$R \text{ (total)} = \frac{R_1 \times R_2}{R_1 + R_2}$$

Thus, in Fig. 16 the effective resistance of R_1 and R_2 is:

$$R \text{ (total)} = \frac{2 \times 4}{2 + 4} = \frac{8}{6} \text{ or } 1.33 \text{ ohms.}$$

In a circuit containing more than two parallel resistors as in Fig. 17 the easiest way to determine the total circuit resistance is as follows: first, assume

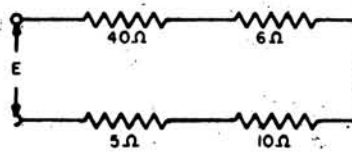


Fig. 15. Resistance in series is added. As far as voltage applied and current flow is concerned the individual resistors are only one.

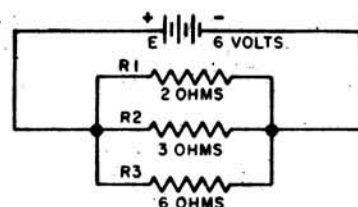


Fig. 18. Series-parallel circuit is not really difficult. Add R_2 and R_3 algebraically. Add effective resistance to R_1 for total resistance.

that a 6-volt battery is connected across the resistor network. Pick a value that will make your computations simple. Then determine the current flowing through each of the resistors using Ohm's law:

$$I = \frac{E}{R_1} = \frac{6}{2} = 3 \text{ amperes}$$

$$I = \frac{E}{R_2} = \frac{6}{3} = 2 \text{ amperes}$$

$$I = \frac{E}{R_3} = \frac{6}{6} = 1 \text{ ampere}$$

Next, add the individual currents flowing through the circuit:

$$2 \text{ amperes} + 3 \text{ amperes} + 1 \text{ ampere} \\ I = 6 \text{ amperes}$$

Inserting this 6 amperes in Ohm's law, the total circuit resistance is found to be:

$$R = \frac{6}{6} = 1 \text{ ohm}$$

The combined equation for determining the total resistance of n number of resistances would be:

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \dots + \frac{1}{R_n}$$

Quite often an electronic circuit will contain a combination of series and

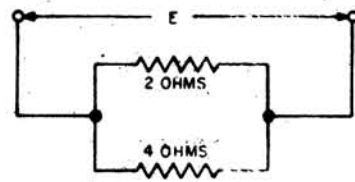


Fig. 16. Resistors in parallel are added algebraically—the result will always be a value less than that of the lowest in the circuit.

parallel resistances as in Fig. 18. To solve this type of problem, first determine the combined resistance of R2 and R3:

$$R \text{ (total)} = \frac{6 \times 12}{6 + 12} = \frac{72}{18} = 4 \text{ ohms}$$

This total value of R2 and R3 may be considered a single resistance which is in series with R1, and forms a simple series circuit. This simple series circuit is solved as follows:

$$R \text{ (total)} = 6 + 4 \text{ or a total of } 10 \text{ ohms.}$$

Power. The amount of work done by electricity is termed the *watt* and one watt is equal to one volt multiplied by one ampere. This may be expressed as: $P = E \times I$ where E = voltage in volts, I = the current in amperes. Also:

$$P = \frac{E^2}{R} \text{ and } P = I^2 R$$

As an example, assume that a toaster draws 5 amperes at an applied voltage of 115 volts. Its wattage would then be:

$$P = 115 \times 5 \text{ or } 575 \text{ watts.}$$

Magnetism and the Electron. The atom and a concept of its structure were a necessary preface to our discussion of basic electricity. By the same token, both are necessary to understanding basic magnetism.

As we've mentioned, electrons are in continual motion about the nucleus. The orbit is, in fact, a small loop of current and has a magnetic field that's associated with a current loop. In addition, experimental and theoretical investigation seems to indicate that the electron itself has a spin. Each electron, having its own axis, is a spinning sphere of electric charge. *Electron spin*, like the quantum and wave theories of light, is not so much a literal interpretation of a phenomenon as a useful concept that holds water when applied to the phenomenon of magnetism.

When the electron spins, the charge that is in motion produces a magnetic field. And, to briefly state the electronic explanation of magnetism, it seems that the magnetic properties of matter can be attributed to the orbital and spinning motion of the electrons comprising the atoms of the matter.

Millennia of Magnetism. Some of the basic principles and effects of magnetism have been known for centuries. The Greeks are credited as the ones who first discovered magnetism. They noted that a certain type of rock had

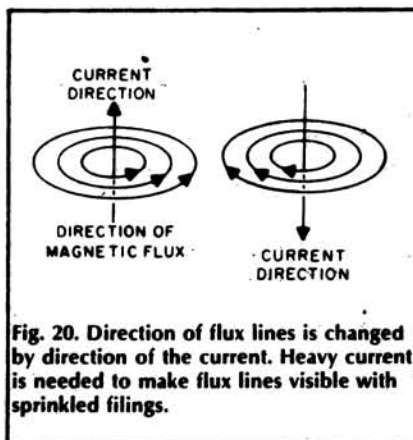


Fig. 20. Direction of flux lines is changed by direction of the current. Heavy current is needed to make flux lines visible with sprinkled filings.

the ability of attracting iron. Later, the Chinese noted that an elongated piece of this rock had the useful property of always pointing in a north-south direction when suspended by a string. This was the beginning of our compass.

This strange stone which intrigued people over the centuries is actually a form of iron ore known as magnetite. Not all magnetite shows magnetic properties. Another name for the magnetic variety of magnetite is lodestone—the term lodestone being derived from two separate words, lode and stone. The term “lode” stands for guide, hence lodestone means “guide stone.”

All magnets, whether natural or man-made, possess magnetic poles, which are commonly known as the magnet's north and south poles. As is the case of the electrical charges (which we studied earlier) between unlike magnetic poles and repulsion between like poles, it has been found that this magnetic attraction and repulsion force varies inversely as the square of the distance from the magnetic poles.

The Magnetic Field. We all know how a magnet exerts a force of attraction on a piece of magnetic material such as iron or steel. Also, when the north poles of two magnets are brought

close together, they will try to repel each other, while there will be attraction between the north and south poles of two magnets. Although it is not clearly understood just what this force of magnetic attraction and repulsion is, it is convenient to visualize magnetic lines of force which extend outward from one magnetic pole to the other as illustrated in Fig. 19.

Permeability. Magnetic lines of force can pass through various materials with varying ease. Iron and steel, for example, offer little resistance to magnetic lines of force. It is because of this that these materials are so readily attracted by magnets. On the other hand, materials such as wood, aluminum and brass do not concentrate or encourage the passage of magnetic lines of force, and as a consequence are not attracted by magnets.

The amount of attraction a material offers to magnetic lines of force is known as its permeability. Iron and steel, for example, possess high permeability since they offer little resistance to magnetic lines of force. Nonmagnetic materials have low permeability. For practical purposes, we can say that reluctance is to magnetic lines of force what resistance is to an electrical current.

Electromagnetism. Any electrical conductor through which flows an electrical current will generate a magnetic field about it which is perpendicular to its axis as shown in Fig. 20. The direction of this field is dependent upon the direction of current flow, and the magnetic field strength proportional to the current strength. If this current-carrying conductor is wound into a coil, forming a solenoid, the magnetic field will be increased by each individual turn that is added. If an iron core is inserted in this current-carrying coil, the generated field will be increased still further. This is because the lines of force are concentrated within the iron core which has considerably less reluctance than the surrounding air.

Self Induction. As mentioned a short while ago, a magnetic field is built up around a coil at the application of current through the coil. As this field is building up, its moving lines of flux will cut the turns of the coil inducing a counter-electromotive force or counter-EMF which opposes the current flowing into the coil.

The amount of counter-EMF generated depends upon the rate of change in amplitude of the applied current as well as the inductance of the coil. This

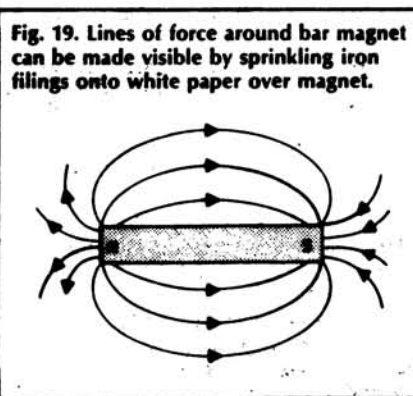


Fig. 19. Lines of force around bar magnet can be made visible by sprinkling iron filings onto white paper over magnet.

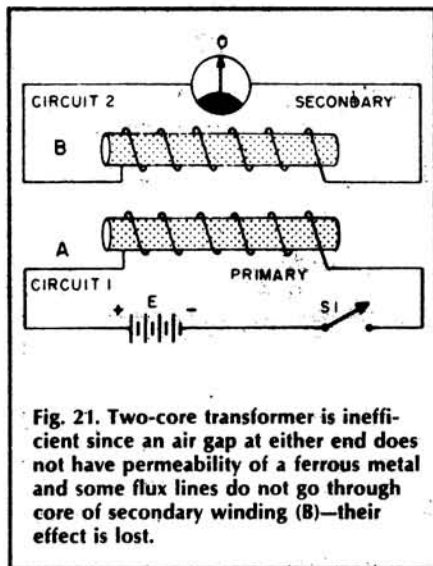


Fig. 21. Two-core transformer is inefficient since an air gap at either end does not have permeability of a ferrous metal and some flux lines do not go through core of secondary winding (B)—their effect is lost.

value of inductance is dependent upon the number of turns in the coil; a coil with many turns will have greater inductance than a coil with few turns. Also, if an iron core is inserted into the coil, the inductance of the coil will increase sharply. The unit of inductance is known as the *henry*.

The Transformer. One of the most important and widely used applications of magnetic induction is the transformer. Transformers find the major application in stepping up or down voltage and current in countless applications.

Fig. 22 shows the basic construction of a typical transformer. While two separate windings are shown here, some transformers can have as many as five or six windings.

A transformer consists of two or more separate windings, electrically insulated from each other. One winding, known as the primary winding, is fed from a source of alternating current.

The alternating currents flowing through the primary induce a current in the secondary winding by virtue of magnetic induction. The transformer core is constructed from a relatively high permeability material such as iron which

readily conducts magnetic flux between the primary winding and secondary winding.

The alternating current flowing in the primary of the transformer produces a variation in the magnetic flux circulation in the transformer core which tends to oppose the current flowing in the primary winding by virtue of self-induction. The counter-EMF is just about equal to the voltage applied to

The magnetic power of a multi-turn current-carrying coil through which a core is inserted is proportional to the current flowing through the coil as well as the number of turns in the coil. The current through the coil is termed *ampere turns*. As an example, if a coil consisting of 200 turns is carrying 2 amperes, its ampere turns equal:

$$\text{Ampere turns} = 200 \text{ turns} \times 2 \text{ amperes or } 400 \text{ ampere turns}$$

Similarly a coil of 100 turns through which a current of four amperes flows also has 400 ampere turns.

Electromagnetic Induction. We saw earlier how a current-carrying conductor will generate a magnetic field which is perpendicular to the conductor's axis. Conversely, a current will be induced in a conductor when the conductor is passed through a magnetic field. The strength of this induced current is proportional to both the speed at which it passes through the field and the strength of the field. One of the basic laws pertaining to electromagnetic induction is Lenz's law which states: "The magnetic action of an induced current is of such a direction as to resist the motion by which it is produced."

Fig. 21 illustrates two coils, A and B, which are placed in close proximity to each other. Coil A is connected in series with a switch and battery so that a current may be sent through it when the switch is closed, and coil B is connected with a current-indicating DC meter. When the switch is closed, current will flow through coil A, causing a magnetic field to be built up around it. In the brief instant that the field is

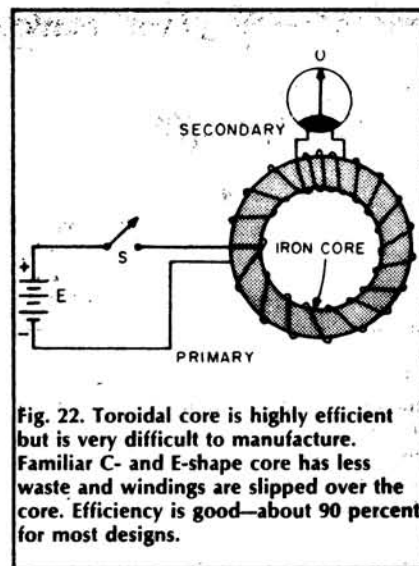


Fig. 22. Toroidal core is highly efficient but is very difficult to manufacture. Familiar C- and E-shape core has less waste and windings are slipped over the core. Efficiency is good—about 90 percent for most designs.

building up to maximum, it will "cut" the turns of coil B, inducing a current in it, as indicated by a momentary flick of the indicating meter. When the switch is opened, breaking the current flow through coil A, the field around coil A will collapse, and in so doing will again induce a current in coil B. This time, however, the flow of current will be in the opposite direction. The meter will now flick in the direction opposite to when the switch was closed. The important thing to remember is that the conductor must be in motion with respect to the magnetic field or vice versa in order to induce a current flow. You can perform this simple experiment using two coils made of bell wire wrapped around large nails, a few dry cells in series, and a DC center scale meter.

Theory and Practice. We've come a long way from our initial discussion of the atom and its importance for an understanding of electricity and magnetism. And there's still a long way to travel to understand all about the subatomic nucleus and its satellites. But, we move ahead by mixing theory with practice—so, put your new knowledge to work in a project or two! ■

Prefixes and Exponents

□ Anyone who's dipped his little toe into electronics is certain to have run across such terms as *microFarad*, *milli-Henry*, and *milliAmpere*—not to mention *megaHertz*, *megOhm*, and *kilo-Hertz*. The prefixes here, *micro-*, *milli-*, *mega-*, and *kilo-*, are an important part of the electronic vocabulary. It follows, then, that anyone who wants to be proficient in electronics will have to de-

velop skill in understanding and using them.

These prefixes are used to change the value of an electronic unit of measure. For example, if you see a resistor with the familiar brown/black/green color code, you *could* call it a 1,000,000-ohm resistor. The thing is, it's usually less awkward to call it a 1-megohm resistor. Putting the prefix *meg-* or *mega-* before

the Ohm inflates the value of the unit, Ohm, by 1,000,000 times.

Similarly, one kiloVolt is recognizable as 1,000 Volts, and one kiloHertz as 1,000 Hertz, and so on. These prefixes are usually so automatic with electronics aficionados that they will invariably refer to a millionaire as a guy who has one megabuck!

The Debit Side. At the other end of

the scale, the *milli-* and *micro-* prefixes are useful for shrinking units. A Farad, for example, is too big a unit to use in everyday electronics. In dealing with the real-life capacitors (the kind you solder into circuits), we normally use a basic unit of *one-millionth* of a Farad—a *microFarad*. The prefix *micro-* cuts up a unit into a million tiny slices, enabling us to use one such slice as a convenient-sized unit. A *microAmpere*, similarly, is a millionth of an Ampere; a *microVolt*, one millionth of a Volt.

If you need larger slices, the *milli-* prefix is available, which provides a unit only one-thousandth the size of the basic unit. A *milliAmpere*, for example, is a thousandth of an Ampere; that is, it takes 1000 mA (*milliAmperes*) to equal 1 Ampere.

To handle these tiny slices of units, it's wise to spend a few minutes learning *scientific notation*, which is designed to make it easy to handle very large and very small numbers. Once you've mastered this technique, you can manipulate all the various-sized units of electronics as easily as you can add two and two!

Take, for example, the familiar *kiloHertz* (known at one time as the *kilocycle*). A broadcasting station operating at 840 kHz (*kiloHertz*) in the broadcasting band is radiating 840,000 cycles of RF energy every second. To change from 840 kHz to 840,000 Hz, you can think of the "kilo-" as being replaced by "x 1000", thus:

840	kilo	Hertz
840	× 1000	Hertz
840,000		Hertz

But you can also write "1000" as "10 x 10 x 10". And you can write "10 x 10 x 10" as "10³". (Ten to the third power, or ten cubed.) As we develop these ideas further, you will see how you can greatly simplify your future work in electronics by thinking of the prefix "kilo-" as being replaceable by "x 10³", thus:

$$840 \text{ kiloHertz} = 840 \times 10^3 \text{ Hertz}$$

Similarly, a 6.8 megohm resistor, measured on an ohmmeter, will indicate 6,800,000 ohms. In this case, the prefix "meg-" can be replaced by "x 1,000,000":

6.8	meg	Ohms
6.8	× 1,000,000	Ohms
6,800,000		Ohms

But you can write "1,000,000" as "10 x 10 x 10 x 10 x 10 x 10" (six of 'em; count 'em), which is 10⁶. Thus, you

ELECTRONIC PREFIXES AND THEIR MEANINGS

Prefix	Pronunciation	Symbol	Exponent	Example
tera-	TEHR-uh	T	10 ¹²	Frequency of infrared light is approx. 1 <i>teraHertz</i>
giga-	GIG-uh	G	10 ⁹	Frequency of TV channel 82 is approx. 1 <i>gigaHertz</i>
mega-	MEG-uh	M	10 ⁶	Frequency of a typical medium wave broadcast station is 1 <i>megaHertz</i>
kilo-	KILL-oh	k	10 ³	Top note on a piano is approx. 4 <i>kiloHertz</i>
hecto-	HEK-toh	h	10 ²	(not often used in electronics)
deka-	DEK-uh	da	10 ¹	(not often used in electronics)
deci-	DESS-ih	d	10 ⁻¹	A decibel is 1/10th bel
centi-	SENT-ih	c	10 ⁻²	Wavelength of TV channel 82 is approx. 30 centimeters
milli-	MILL-ee	m	10 ⁻³	Collector current of a typical small transistor is approx. 1 <i>milliAmpere</i>
micro-	MY-kroh	μ	10 ⁻⁶	Base current of a typical small transistor is approx. 20 <i>microAmperes</i>
nano-	NAN-oh	n	10 ⁻⁹	Time for a radio wave to travel 1 foot is approx. 1 <i>nanosecond</i>
pico-	PY-koh	p	10 ⁻¹²	Collector-to-base capacity of a good high-frequency transistor is approx. 1 <i>picoFarad</i>
femto-	FEM-toh	f	10 ⁻¹⁵	Resistance of 6 <i>microinches</i> of 0000 gauge wire is approx. 1 <i>femtoOhm</i>
atto-	AT-toh	a	10 ⁻¹⁸	6 electrons per second is 1 <i>atto-Ampere</i>

should learn to mentally replace "meg-" with x 10⁶, so that 6.8 megOhms becomes a 6.8 x 10⁶ Ohms. The 6 is called an *exponent*, and shows how many 10s are multiplied together.

The Minus Crowd. What about the "milli-" and "micro-" prefixes? "Milli-", we've said, is one-thousandth; in a way, it is the opposite of the "kilo-" prefix. Make a mental note, then, that milli- can be replaced with "10⁻³" (read as "ten to the *minus* three power"), which is 1/10 x 1/10 x 1/10 = 1/1000. Similarly, the "micro-" prefix can be considered as the opposite of "meg-", and replaced by 10⁻⁶.

The beauty of this approach appears when you are faced with a practical problem, such as, "if 1.2 *milliAmperes* flows through 3.3 megOhms, what voltage appears across the resistor?" From our knowledge of Ohm's law, we know that E = IR; that is, to get Volts (E) we multiply current (I) times resistance (R). Without the aid of scientific notation, the problem is to multiply 0.0012

Amperes by 3,300,000 Ohms, which is rather awkward to carry out. The same problem, however, is very easy in scientific notation, as can be seen below:

$$\begin{array}{r} 1.2 \times 10^3 \\ 3.3 \times 10^6 \\ \hline 3.96 \times 10^3 \end{array}$$

The answer is 3.96 x 10³ Volts, or 3.96 kiloVolts. We obtained the answer by multiplying 1.2 x 3.3 to get 3.96, and adding the -3 exponent to the 6 exponent to get 3 for the exponent of the answer. The advantage of scientific notation is that the largeness and smallness of the numbers involved is indicated by numbers like 10⁶ and 10⁻³, and the largeness or smallness of the answer is found by *adding* the 6 and the -3.

What about a division problem? For the sake of a good illustrative example, consider the unlikely problem of finding the current when 4.8 megaVolts is applied across 2 kilOhms. The problem is written as:

$$I = \frac{E}{R}$$

$$= \frac{4.8 \text{ megaVolts}}{2 \text{ kilOhms}} = \frac{4.8 \times 10^6 \text{ Volts}}{2.0 \times 10^3 \text{ Ohms}}$$

$$4.8 \div 2 = 2.4$$

$$2.4 \times 10^3 \text{ Amperes} = 2.4 \text{ kiloAmperes}$$

In division, then finding the size of the answer becomes a subtraction problem, in which the exponent representing the size of the divisor ("bottom" number) is subtracted from the exponent representing the size of the dividend ("top" number).

A more practical division problem answers the question, "What current flows when 5 Volts is applied across 2.5 kilOhms?"

$$I = \frac{E}{R} = \frac{5 \text{ Volts}}{2.5 \text{ kilOhms}}$$

$$= \frac{5.0 \times 10^0}{2.5 \times 10^3} = \frac{5.0 \times 10^{(0-3)}}{2.5}$$

$$= 2.0 \times 10^{-3} \text{ Amperes}$$

$$= 2.0 \text{ milliAmperes}$$

Note that it's perfectly legal to use 10^0 (ten to the zero power) to indicate a unit that has no prefix—in other words, one of anything.

For the Solving. Here are a few more problems:

1. The inductive reactance of a coil is given by

$$X_L = 2\pi fL$$

What is the reactance of a coil whose inductance $L = 22$ milliHenries, when an alternating current of frequency $f = 1.5$ megaHertz is applied to it?

$$X_L = 2 \times \pi \times (1.5 \times 10^6) \times (22 \times 10^{-3})$$

$$= 207.24 \times 10^3 \text{ Ohms}$$

$$= 207.24 \text{ kilOhms}$$

2. An oscillator is connected to a wavelength-measuring apparatus, and the wavelength of its oscillations is determined to be 2.1 meters. What is the frequency of the oscillator?

$$F = \frac{\text{speed of light}}{\text{wavelength}}$$

$$= \frac{3.0 \times 10^8 \text{ meters per second}}{\text{wavelength}}$$

$$= \frac{3.0 \times 10^8}{2.1 \times 10^0} = 1.4286 \times 10^8 \text{ Hertz}$$

We wish this answer had come out with a "10⁶", instead of a "10⁸", because we can convert 10⁶ Hertz directly to megaHertz. However, we can change the answer to 10⁶, by shifting the decimal point of the 1.4286. Remember this rule: To lower the exponent, shift the decimal point to the right. (Of course, the opposite rule is also true.) Since we wish to lower the exponent by 2, we must shift the decimal point to the right by two places:

$$142.86 \times 10^6 \text{ Hertz} = 142.86 \text{ megaHertz}$$

3. A 3.3 microfarad capacitor is being charged from a 20-volt battery through a 6.8-kilohm resistor. It charges to half the battery voltage in a time given by $T = 0.69RC$

For the particular values given in the problem, what is the time taken to charge to half the battery voltage?

$$T = 0.69 \times (6.8 \times 10^3) \times (3.3 \times 10^{-6})$$

$$= 15.4 \text{ milliseconds}$$

4. A 365-pF variable capacitor and a 2-uH coil are found collecting dust in your junk box. You decide you might like to incorporate them into a radio but you need to know the resonant frequency of this inductive/capacitive circuit. You apply the formula:

$$f = \frac{1}{2\pi\sqrt{LC}}$$

Since $C = 365$ -pF or 365×10^{-12} farads and $L = 2$ -uH or 2×10^{-6} henrys we can use these numbers, the formula and our new knowledge of exponents to determine the frequency.

$$f = \frac{1}{2\pi\sqrt{(2 \times 10^{-6}) \times (365 \times 10^{-12})}}$$

$$= 5,894,627.6 \text{ Hertz}$$

$$= 5,894 \text{ kiloHertz}$$

$$= 5.894 \text{ megaHertz}$$

Tera to Atto. Since scientific notation is so potent, you'll probably be interested in the meaning of *all* the prefixes used in the scientific community, not just the four (micro-, milli-, kilo-, and mega-)—that we've discussed so far. Very common in electronics is the *micro-microFarad*, which is $10^{-6} \times 10^{-6}$ Farad, or 10^{-12} Farad. This is more commonly known as the *picoFarad*. Similarly, a thousandth of a microAmpere is $10^{-3} \times 10^{-6}$ Ampere, or 10^{-9} Ampere. This is known as a nanoAmpere. At the other extreme, 1000 megaHertz is called a gigaHertz. See the table of all these prefixes for a rundown of their meanings and pronunciations.

The jargon of electronics which has grown up around their prefixes is just as important as the prefixes themselves. Here are some examples of "jargonized" prefixes as they might appear in speech:

Puff—a picoFarad (from the abbreviation, PF).

Mickey-mike—a micro-microFarad (which is the same as a puff).

Meg—a megohm. Also, less often, a megaHertz.

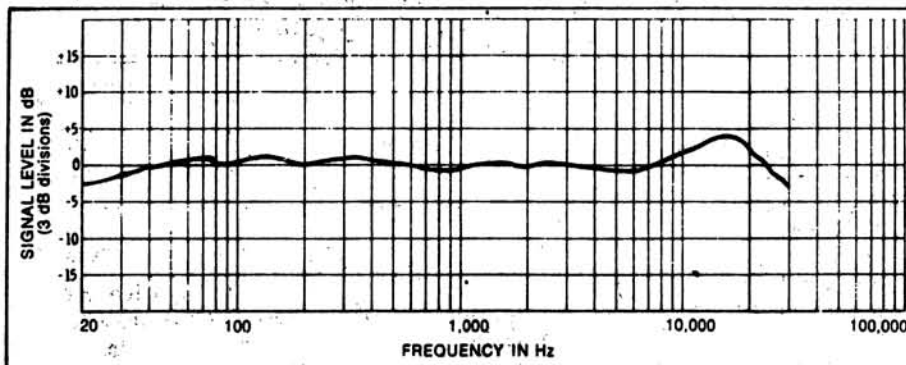
Mill—a milliAmpere.

Megger—a device for measuring megOhms.

dB (pronounced "dee-bee")—a decibel, which is one-tenth of a Bel.

Mike—a microFarad. Also, to measure with a micrometer.

So, if you understand the prefixes and know their corresponding exponents, you'll have command of another set of important tools to help you do practical work in electronics. In addition, you'll be ready for the inevitable wise guy who'll ask if you can tell him the reactance of a 100-puff capacitor at 200 gigaHertz. After calculating the answer in gigaseconds, reply in femto-Ohms!



A common way to deal with measurements that range across many exponents is to use a logarithmic scale. This audio frequency response chart starts at 20 Hz and extends out past 20,000 Hz. Each major division represents a factor of ten. It allows a graphic representation of a broad bandwidth in a small space.

Capacitance Basics

You will learn how capacitance can be used to block direct current (DC) and to pass alternating current (AC) signals. You will also learn how capacitance causes applied AC voltage to lag behind the current in a given circuit, and how capacitance distorts the voltage waveform of pulses. When you have finished you will be familiar with the units we use to measure capacitance and the factors influencing the size of a capacitor.

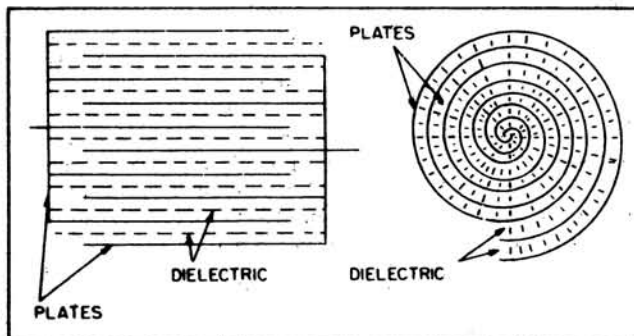
WHAT IS CAPACITANCE?

Capacitance is the property of an electrical circuit that **opposes a change in voltage**. Capacitance has the same reaction to voltage that inductance has to current. This means that if the voltage applied across a circuit is increased, the capacitance will resist that change. If the voltage applied across a circuit is decreased, the capacitance will oppose the decrease and try to maintain the original voltage.

In a DC circuit, capacitance has an effect only when voltage is first applied, and then again when it is removed. **Note that direct current cannot flow through a capacitance.** However, alternating current appears to flow through a capacitance—you will learn how later. Since voltage is constantly changing in AC circuits, capacitance acts at all times to retard these changes.

A basic capacitor is shown in the first diagram. It consists of two conducting metal plates separated by a layer of air or other insulating material, such as paper, glass, mica, oil, etc. The layer is called the **dielectric**.

STACKED AND ROLLED CAPACITOR TYPES



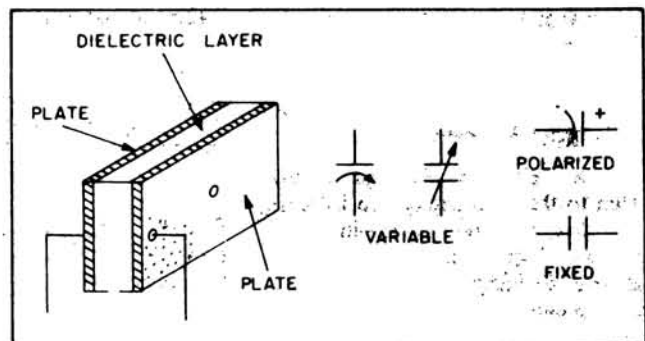
All capacitors have these two plates and a separating layer. In practice, the plates and dielectric are often stacked or even rolled into a compact form. Sometimes the dielectric is a paste or a liquid instead of a solid.

When a capacitor is first connected to a battery, electrons from the negative terminal of the battery flow to the nearest capacitor plate and remain there. They can go no farther, since the opposite plate is separated from the first by an insulating layer. Electrons are attracted from the opposite capacitor plate and flow into the positive terminal of the battery. After this initial move-

ment of electrons, the negative-most plate of the capacitor is filled with all the electrons that the battery voltage can force into it, and the other capacitor plate loses the same number of electrons to the battery's positive terminal. This means that one plate has a negative charge and the other plate has a positive charge—the charge being equal to the battery's potential. No further current flows; the capacitor is "charged."

Positive and negative charges attract each other, so there will be a force between the plates of the capacitor. There is also a voltage between them that is equal to,

CIRCUIT SYMBOLS AND CORRESPONDING PARTS



and which opposes the voltage of the battery.

Because it takes a certain specific number of electrons to fill the negative plate, we say that the capacitor has a certain capacity, or **capacitance**.

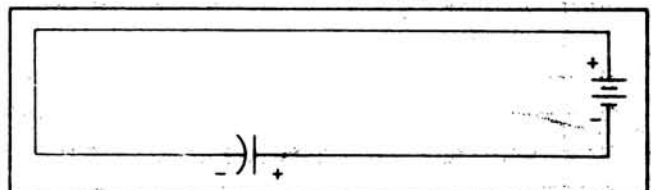
QUESTIONS

- Q1. Name two differences between capacitance and inductance.
- Q2. Draw a circuit diagram of a capacitor connected across the terminals of a battery.
- Q3. Explain what happens when you disconnect the battery terminals from a charged capacitor and place a wire across the leads of the capacitor.

ANSWERS

- A1. Capacitance opposes a change in voltage while inductance opposes changes in current. **Capacitance blocks DC** while inductance does not.
- A2. Your circuit diagram should look like this.

ANSWER TO QUESTION 2



- A3. The **electrons** from the capacitor's negative plate flow through the wire to the positive plate until both plates have the same number of electrons. The **voltage across the plates is then zero.**

CAPACITANCE MEASUREMENTS

The usual written symbol for capacitance is **C**. Capacitance is measured in **farads**. The amount of capacitance in a capacitor is the quantity of electrical charges (measured in coulombs) which must be moved from one plate to the other in order to create a potential difference of 1 volt between plates. The number of coulombs transferred is called the **charge**.

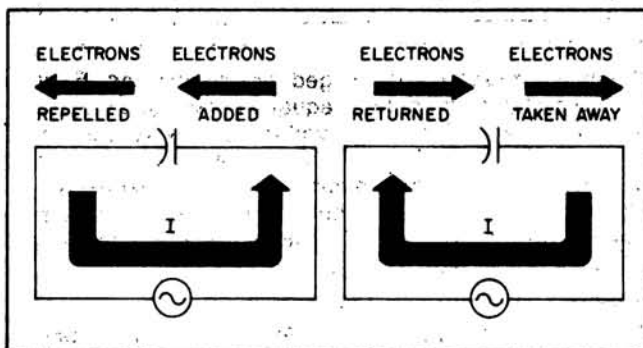
One farad is the capacitance in which a charge of 1 coulomb produces a difference of 1 volt between the plates. The larger the area of a capacitor's plates, and the closer these plates are to each other, the more charge (current) the capacitor will hold with the same voltage applied across the plates.

Capacitance values are usually specified in microfarads (millionths of a farad, abbreviated **mfd** or μF) or in picofarads (millionths of a microfarad, abbreviated **pF**).

HOW DOES CAPACITANCE AFFECT AC?

Although current cannot flow through a capacitor, an AC current appears to do just that. The reason lies in the nature of capacitance. If the voltage across the plates is continuously varied, the number of electrons on the plates varies.

AC THROUGH A CAPACITOR



Increasing the number of electrons on one plate of a capacitor repels electrons from the other plate. Decreasing the number of electrons on the first plate allows electrons to be attracted back to the other plate.

An AC current can, in effect, get across the dielectric. Since the voltage is alternating, it causes a corresponding varying current to flow between one side of the capacitor to the other side. In other words, **voltage changes** appear to be transmitted across the dielectric.

If a capacitor has the same voltage as the applied voltage, no current will flow to or from it. If the applied voltage changes, the capacitor voltage will no longer equal the applied voltage. Current will flow, trying to equalize the two potential sources.

In a circuit this means that if an AC sine-wave voltage is applied across a capacitor, an AC sine-wave current will appear on the opposite side, even though no electrons flow through the dielectric layer.

QUESTIONS

- Q4. The capacitance of a capacitor is measured in _____.
- Q5. A millionth of a farad is called a _____ and is abbreviated as _____ or _____.
- Q6. A _____ is a millionth of a microfarad and is abbreviated _____.
- Q7. Current will flow from one plate of a capacitor to

the other plate only when _____ is changing.

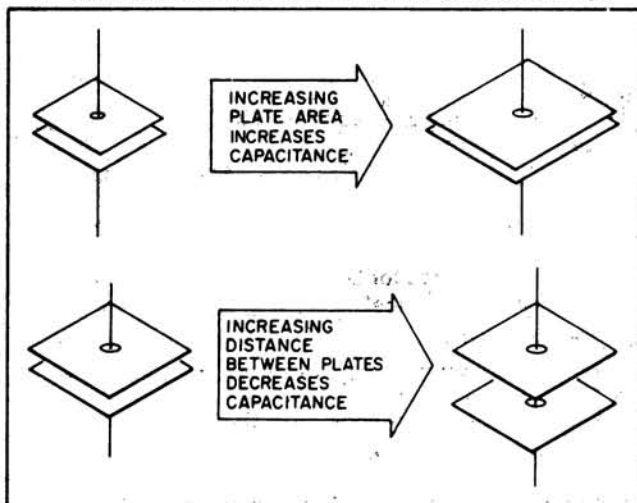
ANSWERS

- A4. The capacitance of a capacitor is measured in **farads**.
- A5. A millionth of farad is called a **microfarad** and is abbreviated as **mfd** or μF .
- A6. A picofarad is a millionth of a microfarad and is abbreviated **pF**.
- A7. Current will flow through a capacitor only when **voltage** is changing.

FACTORS AFFECTING CAPACITANCE VALUE

The amount of electrical charge that can be stored in a capacitor (the number of electrons that can be placed on the plate) varies with the **area** of the plates. Consequently, capacitance varies directly with area—if the

PHYSICAL FACTORS AFFECTING CAPACITANCE



area is doubled, the capacitance is doubled. When the area is doubled, or twice as many plates are connected in parallel, there is twice as much area to store electrons. Therefore the capacitance is twice as great.

Capacitance can also be increased by placing the plates closer together. When the plates are closer the attraction between the negative charges on one side and the positive charges on the other side is greater. It is, of course, necessary to keep the plates sufficiently separated so that the charge does not jump through the dielectric, possibly damaging the capacitor.

Higher values of capacitance can be obtained by using an insulating material (dielectric) other than air. In this way the plates can be placed closer together.

Dielectrics such as mica, glass, oil, and **mylar** are a few of the materials that can withstand a high electric potential without breaking down. This property is called **dielectric constant**. The higher the dielectric constant, the better its ability to retain its insulating characteristics under unusual operating conditions. Air has a constant of 1, glass about 5, and mica 2.5 to 6.6.

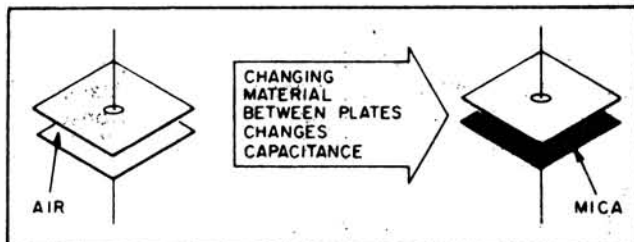
Besides allowing the plates to be placed closer together, a dielectric has another effect on capacitance. Dielectric material contains a large number of electrons and other carriers of electrical charge. Although electrons cannot flow as in a conductor, they are held rather loosely in the structure and can move slightly. The distortion of the structure of the dielectric, which is caused

Understanding Electronics

by charging the capacitor, has a large effect on the forces of attraction and repulsion that aid or oppose the flow of the electrons. This factor has a substantial effect on capacitance.

When materials such as mica or glass are used as the dielectric, the capacitors have a much higher value than the same size units with an air dielectric.

DIELECTRIC FACTORS AFFECTING CAPACITANCE



QUESTIONS

- Q8. How does a mica capacitor differ from an air capacitor of the same physical size?
- Q9. What are three factors that affect the capacitance of a capacitor?
- Q10. A screw-type variable capacitor is made with an adjusting screw that is used to vary the distance between the capacitor plates. How would you increase its capacitance?

ANSWERS

- A8. A mica capacitor has a **higher capacitance** than an air capacitor of the same physical size.
- A9. The capacitance of a capacitor depends on these three factors: the **area** of the plates, the **spacing** between the plates, and the nature of the **dielectric material**.
- A10. Tightening the screw moves the plates closer together and **increases capacitance**. Loosening the screw **decreases the capacitance**.

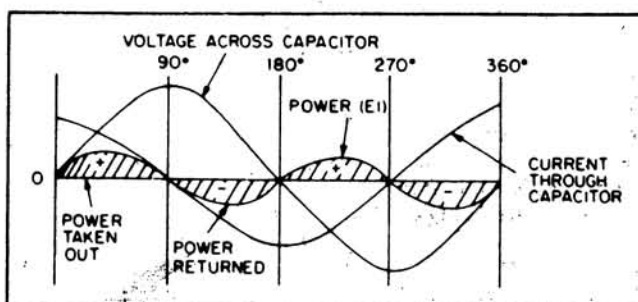
POWER

A perfect capacitor consumes no power. During the sine-wave cycle, the capacitor takes energy out of the circuit and stores it in the form of an electric field during a quarter cycle. The capacitor returns it to the circuit in the next quarter cycle. **Energy is borrowed, but it is returned later.**

If the product of E times I is taken at every instant of the cycle, the power waveform will show that energy is taken out and returned in alternate quarter cycles.

To find the amount of energy (in **coulombs**) stored in

ENERGY STORAGE CYCLE



a capacitor, multiply the capacity in farads by the applied voltage. In a circuit containing only pure capacitance, it makes no difference how long the voltage is applied—the same amount of energy will always be stored at a given voltage.

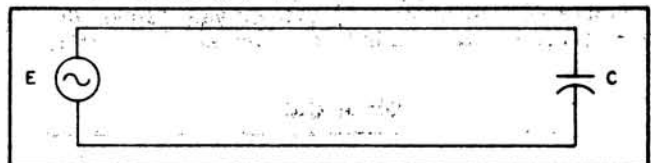
CAPACITIVE REACTANCE

Like inductance, capacitance has a reactance—an opposition to the flow of AC. But capacitive reactance **decreases** as frequency increases.

Suppose a capacitor is connected in series with an alternating voltage source. There is no resistance present at all in the circuit.

Because the circuit below contains no resistance, the voltage across the capacitor will be the same value as the source voltage at every instant.

RESISTANCE-FREE CIRCUIT



When a capacitor is charged up to voltage E , it stores an amount of energy equal to the capacitance times the voltage. If the peak voltage of the AC source is E , the capacitor will have stored a particular amount of energy every time the voltage sine wave reaches its peak, and again stores that amount whenever the voltage reaches its negative peak. The energy depends only on capacitance and peak voltage.

QUESTION

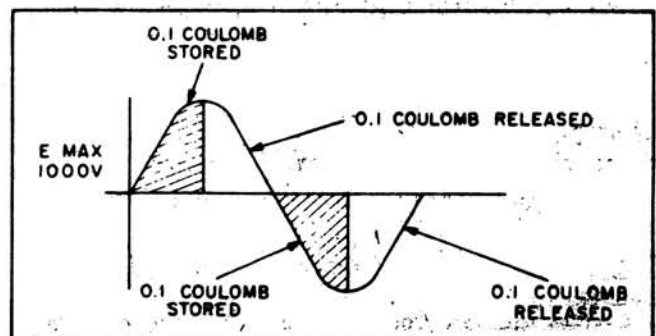
- Q11. How much energy will be stored in a 100-mfd capacitor in the first quarter cycle of an applied AC voltage of 1,000 volts maximum?

ANSWER

- A11. 1,000 volts $\times$.0001 farad = 0.1 coulomb

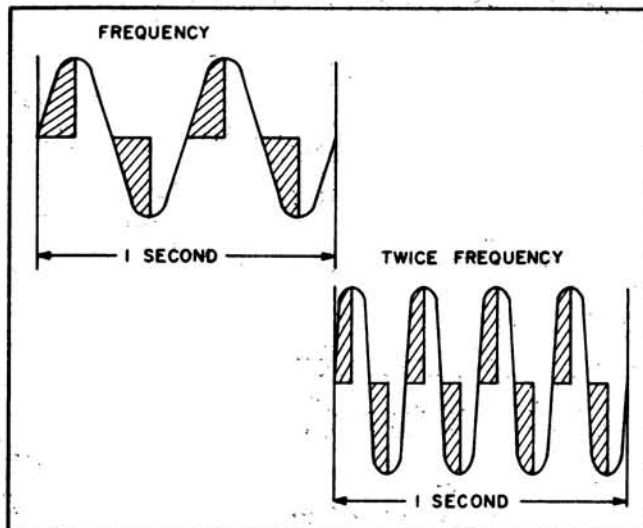
What happens when the frequency of the power source is doubled? If the peak voltage (E) is unchanged, the capacitor will charge every half cycle to the same amount as before. But it will have to do this twice as fast because the energy is doubled. This means that the same amount of energy must flow into the capacitor in only half the time. And since the voltage is the same, we must have twice the current to supply this same amount of energy.

ENERGY STORAGE POTENTIAL



What does this mean? The frequency was doubled, and this doubled the current flowing into the capacitor. Yet, the input voltage remained the same. A pure capacitance lets twice as much current flow if the frequency is doubled.

POTENTIAL VERSUS FREQUENCY



Capacitive reactance is the opposition that pure capacitance offers to the flow of current. It is expressed in **ohms**, and its symbol is X_c . Capacitive reactance depends on frequency. As the frequency increases, the rate of change of applied voltage increases, and the current flowing also increases. As the frequency is reduced, the rate of change of voltage goes down, and less current flows.

At this point you can more easily see why capacitor current leads the voltage across the capacitor. It is necessary for the capacitor to charge up to the given voltage, and this charging is done by the current. Hence, the charging current will reach its maximum value at the time the charging is going on at the greatest rate; that is, when the rate of change of voltage is the most rapid.

As the capacitor approaches full charge, the voltage rate of change slows down, and the current decreases. When the capacitor is fully charged and its voltage has reached maximum, there is no charging current flowing at all—the current has already dropped to zero at this time. A similar process occurs during discharging. At all times, current leads the voltage by 90° , or one quarter of the cycle. In a steady-state AC situation, when the applied voltage is a sine wave, both voltage and current will be sine waves.

Capacitive reactance depends on frequency. Since it lets more current flow as frequency increases, **capacitive reactance must decrease as the frequency increases.**

Capacitive reactance also depends on the size of the capacitance. As capacitance increases, more current must flow into the capacitor to charge it to the same voltage (since the amount of energy stored equals C times E). As a result, **capacitive reactance decreases when capacitance increases.**

The formula for capacitive reactance is:

$$X_c = \frac{1}{2\pi f C} \text{ ohms}$$

where,

f is the frequency in Hz,

C is the capacitance in farads.

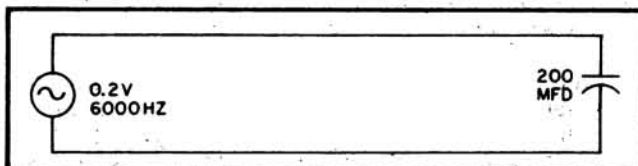
Capacitive reactance can be used in calculating current in a purely capacitive circuit by Ohm's law.

$$I = \frac{E}{X_c}$$

QUESTIONS

Q12. What is X_c if $f = 6,000$ Hz and $C = 200$ mfd?

CIRCUIT FOR QUESTION 12



Q13. What is the current in the circuit?

Q14. What would the current in the above circuit be if the input signal were 0.01 volt at 120 kHz?

ANSWERS

$$\begin{aligned} \text{A12. } X_c &= \frac{1}{2\pi f C} = \\ &= \frac{1}{2 \times 3.14 \times 6,000 \times 200 \times 10^{-6}} \\ &= \frac{1}{7.53} = 0.133 \text{ ohm} \end{aligned}$$

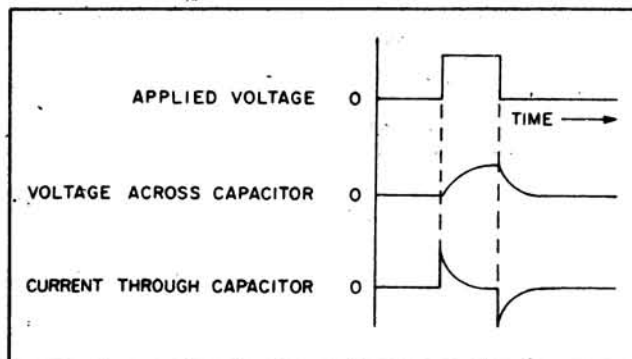
$$\text{A13. } I = \frac{E}{X_c} = \frac{0.2}{0.133} = 1.5 \text{ amps}$$

$$\text{A14. } I = \frac{0.01}{0.0066} = 1.52 \text{ amps}$$

PULSE RESPONSE OF CAPACITANCE

When a sharp pulse, such as a square wave, is applied to a circuit containing capacitance, the capacitance opposes the sudden change of voltage. This results in a rounding off of the sudden voltage rise. Similarly, when the pulse voltage is suddenly decreased, the voltage across the capacitor does not decrease suddenly, but it trails off. Current is greatest when the change of voltage is greatest, so the current waveform will have a peak when the voltage rises suddenly, and another peak (but in the opposite direction) when it drops.

CAPACITANCE VERSUS TIME



Understanding Electronics

There is always some resistance in a practical circuit. By choosing the right values of capacitance and resistance, a circuit can be designed in which the voltage takes a predetermined length of time to reach a certain value. This type of circuit can provide a time delay.

STRAY CAPACITANCE

Capacitive reactance decreases as frequency increases. In communications, pulse, and radar work, where very high frequencies are used, **stray capacitance** can present quite a problem.

In a vacuum tube, an antenna, or a receiver chassis, there are always small capacitances between adjacent conductors and between conductors and nearby objects which are meant to be isolated from each other. With audio and lower radio frequencies these capacitances are not important. But as the frequency increases, the capacitive reactances of these small capacitances decrease. Enough decrease in reactance can actually cause leakage of the signal.

Thus, at high frequencies, placement of wires and components is very important in order to keep the effects of stray capacitance to a minimum.

QUESTIONS

Q15. How does capacitance affect pulses?

Q16. Compare and contrast capacitive reactance and inductive reactance on these points:

1. Effect of an increase in frequency on reactance.
2. Effect of reactance on DC.
3. Effect of phase relations in AC.

Q17. What constant value appears in the formulas for both capacitive and inductive reactance?

ANSWERS

A15. Capacitance rounds off the voltage waveform and produces spikes in the current waveform.

A16. 1. X_C decreases as frequency increases, while X_L increases.

2. X_C blocks DC, while X_L passes DC.

3. Capacitance causes current to lead the applied voltage, while inductance causes it to lag.

A17. 2π appears as a constant in both formulas.

WHAT YOU HAVE LEARNED

1. Capacitance offers opposition to any change in voltage.
2. A basic capacitor consists of metal plates separated by a dielectric.
3. A capacitor stores electrical energy in the form of an electric field as the capacitor charges, and releases this energy when it discharges.
4. Capacitance is a measure of the energy storage capacity of a capacitor. This capacity is measured in farads.
5. A capacitor blocks DC but allows AC to flow.
6. Pure capacitance in a circuit causes current to lead the applied voltage by 90° .
7. The amount of capacitance is determined by the area of the plates, the distance between them, and the dielectric material.
8. A capacitor stores energy and returns it to the circuit.
9. The opposition of capacitance to the flow of AC is called capacitive reactance.
10. The formula for capacitive reactance is:

$$X_C = \frac{1}{2\pi fC}$$

11. Capacitance rounds off the voltage waveform of a pulse.
12. Stray capacitance can cause signal leakage at high frequencies.

SCR Principles

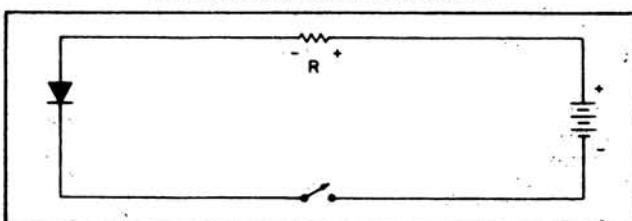
WHAT IS AN SCR?

An SCR, in practice, is basically a **standard rectifier** placed in series with a **switch**. When the switch is closed, the rectifier will conduct in one direction. The difference between the diode/switch combina-

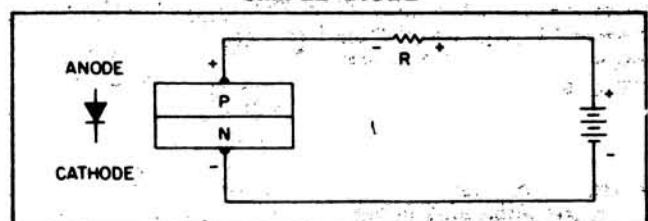
tion and the SCR is that the switch is built right into the SCR itself, and is commonly referred to as the **gate**. More about this later on.

A basic, four-layer silicon device has three junctions. With a battery connected as shown, the device **cannot**

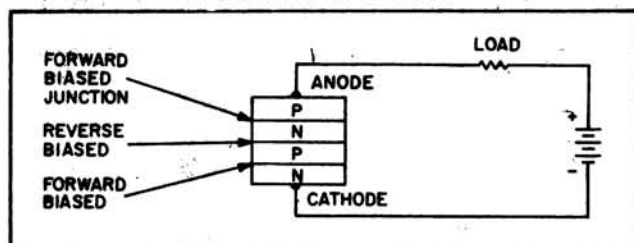
FUNCTIONAL SCR DIAGRAM



SIMPLE DIODE



FOUR-LAYER DEVICE



conduct because no current can flow through the reverse-biased junction in the middle. In this configuration, there are only three ways in which current flow can be induced: the current level can be increased to the point where the junction is simply overwhelmed, light can be directed at the junction until enough electron excitation is created to bridge the reverse-biased junction, or the device can be heated to the point where electron excitation occurs. Of course, these are all clumsy methods.

QUESTIONS

- Q1. The basic component around which an SCR is built is the _____.
- Q2. The SCR's gate functions as a _____.

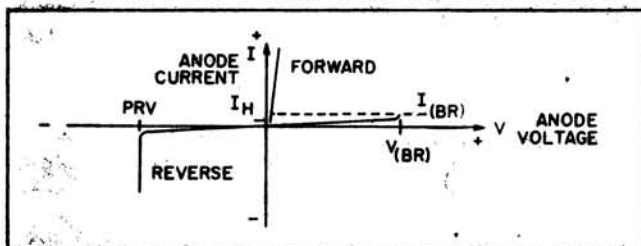
ANSWERS

- A1. diode
A2. switch

TRIGGERING

If we were to plot the voltage and current characteristics of this four-layer device, the result of the **forward voltage breakover** (VBR)—the point at which the SCR begins to conduct—can be observed. Initially, a small current (IBR) is required to trigger the device. In this application, it can be derived from any of the three sources mentioned above. In addition, a **minimum current** level (IH), or holding current, must be present in the external circuit in order to keep the device in a conducting state. Once the device has been triggered, and as long as the holding current is sufficient, no further triggering action is necessary.

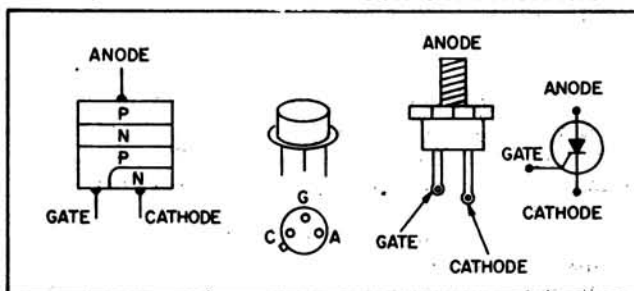
CURRENT AND VOLTAGE BREAKOVER GRAPH



As noted earlier, the physical methods required to trigger the four-layer device are all clumsy and in fact may be destructive to the silicon wafer upon which the device is constructed. The difference between the four-layer device and the SCR, is that the "P" region **nearest the cathode** has a lead attached to it (the **gate**) which, when current is directed into it (IBR), negates the normal reverse bias of the middle junction and allows the SCR to behave like a normal rectifier. Again, once the gate has allowed the cathode to trigger, only a small current level (IH) is required to keep the device in a conducting state. Even with gate current removed,

the device will continue to operate until the holding current in the rest of the circuit falls below the minimum level.

INTERNAL AND EXTERNAL SCR CONFIGURATION



QUESTIONS

- Q3. The _____ is a measure of the point at which an SCR will conduct.
- Q4. It is necessary to apply the _____ to the gate in order for the SCR to conduct.
- Q5. The SCR will cease to conduct only if _____.

ANSWERS

- A3. forward voltage breakover
A4. breakover current (IBR)
A5. the holding current (IH) drops below the threshold level.

SCR OPERATIONAL PROBLEMS

One of the problems that engineers encounter when using the SCR in a switching capacity is **time**. As mentioned earlier, once the SCR has been triggered and current starts to flow from cathode to anode (typically this requires a time period of about 5-microseconds), it will continue to flow through the SCR until and unless the current level drops below the minimum required to hold the SCR open. Suppose, however, that we wish to use the SCR to switch a high-speed DC pulse current. Remember the center junction of the four-layer device that held the reverse (blocking) bias? Because all of the forward voltage flowing through the device must cross this junction, the "P" and "N" regions which comprise the junction actually act as the plates of a **capacitor**, charged to the voltage which is flowing through the SCR.

As a result of this phenomenon, an abrupt application of forward voltage to the junction (as may be supplied by a high-speed pulsed driver or other pulse source) may, through this capacitive effect, cause the SCR to trigger **before** a gate current is ever applied. Typically then, one must allow 50-microseconds to elapse before applying another pulse of forward voltage to the SCR to insure that it has reset itself.

This necessity limits the applications in which the SCR can be used, although newer types of switching transistors can operate at much higher speeds and have, in many cases, obsoleted the SCR in high-speed switching applications.

QUESTIONS

- Q6. It takes _____ for an SCR to reach full conductance level.
- Q7. An SCR can be falsely triggered because the middle junction has a _____ effect.
- Q8. An SCR must be allowed a period of _____ to reset itself.

Understanding Electronics

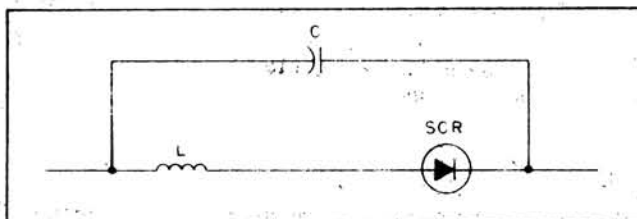
ANSWERS

- A6. 5-microseconds
- A7. capacitive
- A8. 50-microseconds

RADIO FREQUENCY INTERFERENCE

Another problem encountered when incorporating the SCR into a circuit (especially one which operates at radio frequencies, such as digital clocks, ham receivers, and microcomputers) is that the relatively abrupt triggering and conductance process associated with the SCR, and the concurrently swift current level rise through the circuit, can cause RF frequencies to be spuriously generated. These RF signals can be

INDUCTIVE/CAPACITATIVE SHUNT



radiated directly into other parts of the circuit, or can be coupled into the power supply line and travel through it to other parts of the circuit just as easily.

Normal RF shielding precautions can be taken (good grounding, use of shielded cable, etc.), and in addition, an **inductive/capacitive shunt** may be employed to effectively slow down the current rise time across the SCR. Usually, a value of 0.1-uF for the capacitor, and 75-uH (microHenrys) for the inductor will suffice. Be sure, however, that both components are capable of

handling the maximum voltage and current levels which will appear across the SCR normally as a result of the circuit's operation.

QUESTIONS

- Q9. An SCR's switching action may create _____.
- Q10. An _____ will help quiet an SCR.

ANSWERS

- Q9. RF interference
- Q10. inductive/capacitive shunt

SUMMARY

You have learned that the SCR is essentially a rectifier diode with an incorporated switch. The switch is comprised of two extra "P" and "N" layers which form a reverse-biased junction which effectively cuts off the forward flow of current and voltage. The **gate**, which is a lead attached to the "P" junction nearest the cathode, can, when a suitably large current (IBR) is passed along it, negate the reverse bias at the middle junction and allow current to flow. Once the current is fully flowing, removal of the breakover current will not open the gate, and the flow can only be stopped when current in the rest of the circuit drops below the level (IH) necessary to hold the junction open. In addition, you now know that it takes approximately 50-microseconds for the capacitive charge on the "P" and "N" layers adjacent to the middle junction to drop to a level where the junction will not be closed by application of a forward voltage to the cathode. The swift rise of current associated with the closing of the middle junction (just as the closing of a light switch will cause a "pop" to come through an AM radio near it) can create RF interference, and suitable shielding and shunting may be necessary in certain circuit applications to avoid this. ■

Cathode Ray Tubes

WHAT YOU WILL LEARN. When you have finished reading this article you will have learned how the cathode-ray tube, which is the display device for oscilloscopes, as well as for all television receivers, works. In addition, you will know what the differences are between cathode-ray tubes used in 'scopes and those used in TVs.

THE CATHODE-RAY TUBE

The cathode-ray tube (which we'll refer to as CRT from here on) is a large vacuum tube which has three main parts. They are, first, the **electron gun**, which produces a steady stream of electrons and aims them at the large, flat end of the tube, second, **deflecting** devices, which move the electron beam in accordance with the signal to be observed, and third, the chemical coating on the large flat end of the CRT, commonly called the **screen**.

The oscilloscope displays electrical signals on the screen to show what's going on in electrical or electronic circuits. The TV set shows pictures transmitted from the TV station. In both cases the CRT used in

the 'scope or the TV set are almost exactly the same.

The main difference in the picture tube in TV sets and the CRT in scopes today is that the electron beam is moved back and forth in the TV set **electromagnetically**, by coils of copper wire placed around the neck of the tube, while the electron beam in a 'scope is moved about by the changing **electrostatic** voltages between small deflection plates inside the neck of the tube. In fact, the earliest TV picture tubes were electrostatic-deflection CRTs, and during the Korean War, when copper for the magnetic deflection coils was scarce, TV set makers stopped making electromagnetic-deflection picture tubes and went back to the earlier, electrostatically-deflected tubes!

THE CRT IN OSCILLOSCOPES

The oscilloscope is really a not-very-exact **measuring** instrument for voltages and waveforms which also shows what the voltages or signals **look** like. Although it can be pretty exact in its measurement of signals, only the most expensive 'scopes are nearly as precise as even cheap meters, so their main purpose is to

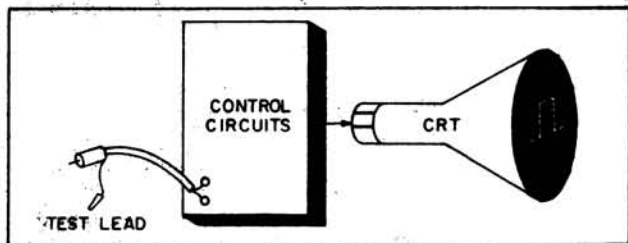
ally to show what the signals look like.

The oscilloscope contains, in addition to the CRT, a power supply which generally provides 2,000-volts or more, and some control circuits which take the signal voltage(s) to be displayed, amplify and otherwise process them and feed them to the CRT for display.

Scope CRTs have a screen usually made of phosphorescent (give off light when struck by electrons) chemicals which create a green display. Some very expensive 'scopes use CRTs with blue, or even purple-emitting phosphor-coated screens. TV sets of course have screens with white light-emitting screens (in black and white sets).

The CRT in oscilloscopes uses electrostatic deflection, with the deflecting voltages applied to the vertical-deflecting plates and to the horizontal deflecting plates, as shown in the diagram.

SIMPLIFIED OSCILLOSCOPE



The signal voltage size is indicated by the *amplitude* (height, up-and-down dimension) of the beam movement on the screen. The *time period* (duration) is shown by the distance the beam travels across the screen horizontally, from left to right.

By relating the time a signal takes to its amplitude (size) and its shape, we can get a very accurate idea of what's going on in most circuits at any desired point.

QUESTIONS

- Q1. A waveform can be described in terms of its vertical and horizontal dimensions. What are these dimensions? amplitude, time.
- Q2. A cathode-ray tube can display a picture on its face, or screen. What causes the picture to appear?
- Q3. An oscilloscope is made up of a cathode-ray tube and a group of control circuits. What is the function of the control circuits?

ANSWERS

- A1. The vertical and horizontal dimensions of a waveform are **amplitude** and **time**.
- A2. The picture on a CRT is developed by a **moving electron beam** that strikes and illuminates a chemical coating on the inside face of the tube.
- A3. The function of the oscilloscope control circuits is to process, amplify, and deliver the **signal** to the CRT.

THE CRT IN TV SETS

The cathode-ray tube is the display device in the television set. The CRT operates by moving a controllable beam of electrons across the inside face of the tube. The number of electrons in the beam is determined by the blacks, grays, and whites of the scene the TV camera is viewing. White is produced by a large number of electrons striking a chemical coating

on the inside of the tube. The electrons cause the coating to give off light. Black is achieved by stopping the electron flow, and shades of gray are obtained by varying the amount of electrons between the amounts required for black and white.

The picture is "painted" on the screen by the narrow electron beam moving back and forth across the tube many times a second. This movement is due to varying magnetic fields produced by a set of horizontal and vertical deflection coils around the CRT's neck.

The principle of putting a picture of a waveform on the screen of an oscilloscope is similar. The movement of the electron beam in the 'scope is controlled *electrostatically* so that the beam traces out the pattern of the waveform being measured. As in the TV tube, the electron beam illuminates a coating on the inside of the tube.

ELECTROSTATIC FIELDS

To understand how the CRT operates, you must know that an electrostatic field is a space in which electric forces act.

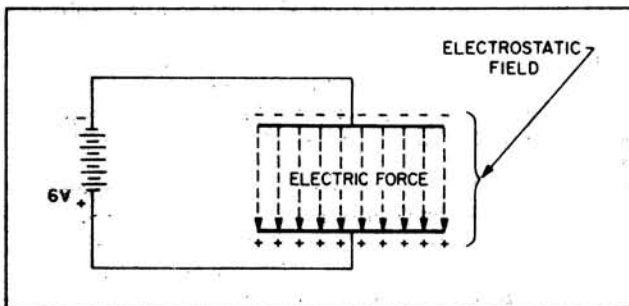
An electrostatic field can be developed between two charged plates. If one plate is negative with respect to the other, the direction of the electric force can be determined.

In the drawing shown, lines of electric force take a direction from negative to positive. This means a negatively-charged body entering the field would be moved downward (from negative to positive). A positively-charged body, however, would be moved upward (positive to negative). (Like charges repel, and unlike charges attract). How do you think an electrostatic field is formed?

An electrostatic field is formed with a voltage source and a pair of metallic plates to hold the charges.

If a 6-volt battery is connected to the plates in the manner shown, the battery will draw electrons from the bottom plate and deposit them on the top plate until the difference in potential between the plates

ELECTROSTATIC FIELDS



equals the battery voltage. The potential of the plate having an excess of electrons will be negative. The other plate, being deficient in electrons, will be positive.

QUESTIONS

- Q4. What is an electrostatic field?
- Q5. What causes an electrostatic field to exist between two metallic plates?

ANSWERS

- A4. An electrostatic field is a region in which **electric forces** are acting.
- A5. An electrostatic field is formed when one plate has

Understanding Electronics

an excess of and the other a deficiency of electrons.

ELECTROSTATIC FORCES BETWEEN CIRCULAR AND TUBULAR PLATES

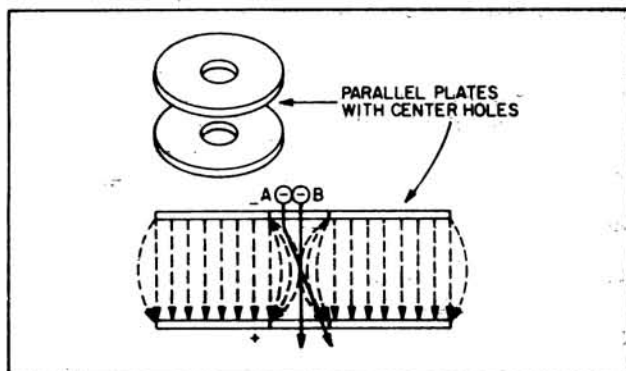
In the drawing, an electrostatic field between two plates having center holes is shown. Observe the curvature of the force lines under the holes.

PRODUCING AN ELECTROSTATIC FIELD

Since its path is parallel to the force lines, electron B will pass straight through the axis (center line) of the holes. Electron A starts in the same direction as electron B. When electron A enters the field, it turns in the direction of the force lines. Just before it leaves the field, it is turned even further and in the direction of the curvature of the force lines.

Suppose a small and a large cylinder, both charged with a positive potential, are placed so the electrons

ELECTROSTATIC LENS I

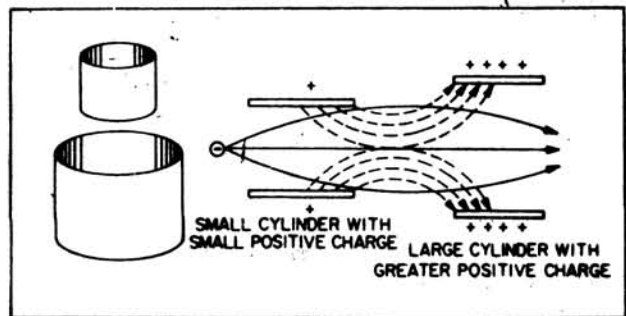


must pass through them. Also suppose the larger cylinder has a more positive charge. The distribution of the lines of force would look like the next illustration.

An electron in the space at the left of the small cylinder would be attracted toward the cylinder by the positive charge. If the electron was travelling along the axis of the cylinder, it would pass through without crossing a line of force. As it approached the larger, more positively charged cylinder, the velocity of the electron would increase.

An electron entering the small cylinder at an angle will cut the lines of force and be turned in their direction as shown by the top and bottom electron paths.

ELECTROSTATIC LENS II



ELECTROSTATIC FOCUS

As it approaches the larger cylinder, the electron

will be accelerated by the higher positive potential. Because of the higher electron velocity, the force lines in the larger cylinder will have a smaller turning effect on the electron. If the difference of potential between the cylinders is adjusted properly, the electrons will unite at a given distance after passing through the second cylinder. The action of the electrons as they pass through the influence of the two cylinders provides a convenient method of focusing the beam.

QUESTIONS

- Q6. As an electron approaches the larger cylinder, the velocity of the electron will
Q7. Why is the above statement true?

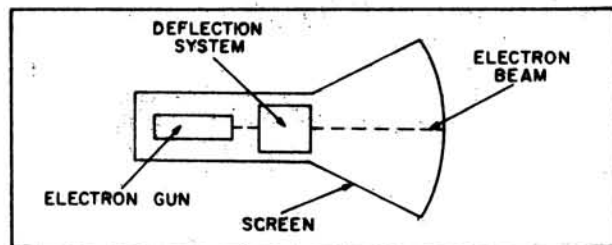
ANSWERS

- A6. As an electron approaches the larger cylinder, the velocity of the electron will **increase**.
A7. The above statement is true because the larger cylinder is **more positively charged**. It will attract the electron with a greater force, thereby increasing the velocity of the electron.

ELECTRON GUN

Cathode-ray tubes used in oscilloscopes consist of an *electron gun*, a *deflection system*, and a *fluorescent screen*. All elements are enclosed in an evacuated container, usually glass. The electron gun generates electrons and focuses them into a narrow beam. The deflection system moves the beam across the screen in the manner desired. The screen is coated with a mate-

BASIC CATHODE-RAY TUBE



rial that glows when struck by the electrons.

An electron gun has a cathode to generate electrons, a grid to control electron flow, and a positive element to accelerate electron movement. The control grid is cylindrical in shape and has a small opening in a baffle at one end. The positive element consists of two cylinders, called anodes. They also contain baffles (or plates) having small holes in their centers. The main purpose of the first anode is to focus the electrons into a narrow beam on the screen. The second anode speeds up the electrons as they pass.

CATHODE AND GRID

The cathode is indirectly heated and emits a cloud of electrons. The control grid is a hollow metal tube placed over the cathode. A small opening is located in the center of a baffle at the end opposite the cathode. The grid is maintained at a negative potential with respect to the cathode.

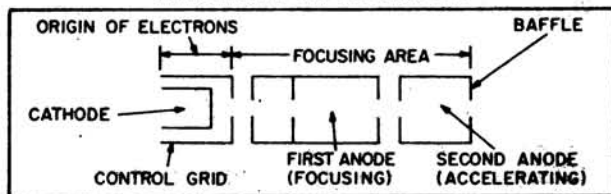
A high positive potential on the anodes pulls electrons through the hole in the grid. Since the grid is near the cathode, it can control the number of elec-

trons that are emitted. As in an ordinary vacuum tube, the negative voltage of the grid can be changed to vary electron flow or stop it completely. The brightness of the image on the fluorescent screen is determined by the number of electrons striking the screen. Intensity (brightness) can, therefore, be controlled by the voltage on the control grid.

FOCUS CONTROL

Focusing is accomplished by controlling the electrostatic fields that exist between the grid and first anode and between the first and second anodes. Study the diagram. See if you can determine the paths of electrons through the gun.

DIAGRAM FOR Q8 and Q9



QUESTIONS

- Q8. Which element controls the number of electrons striking the screen in the drawing titled *electrostatic fields*?
- Q9. Which element controls the focus of the beam?

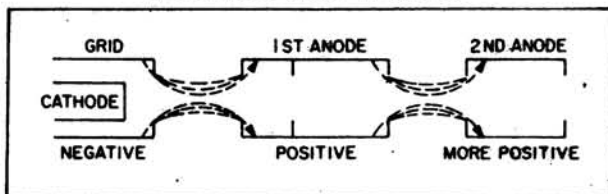
ANSWERS

- A8. The control grid controls the number of electrons striking the screen.
- A9. The first anode controls the focus of the beam.

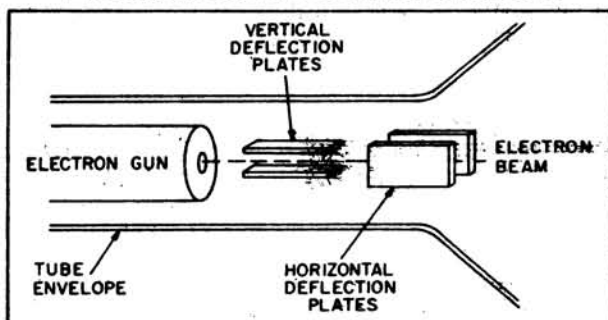
ELECTROSTATIC LENSES

The next diagram shows electrons moving through the gun. The electrostatic field areas are often referred to as *lenses*. The first electrostatic lens causes the electrons to cross at a focal point within the field. The second lens bends the spreading streams and returns them to a new focal point.

ELECTRON GUN "LENSES"



ELECTRON GUN AND BEAM FORMATION



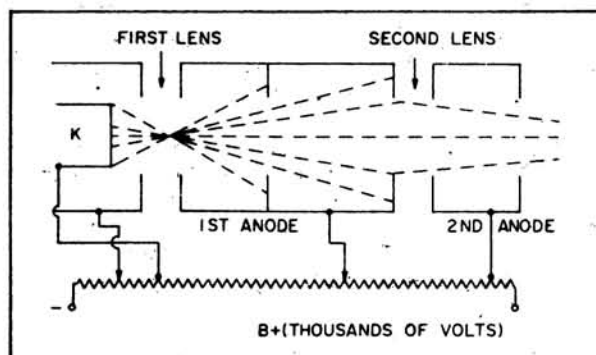
The diagram also shows the voltage relationships on the electron-gun elements. The cathode is at a fixed positive voltage with respect to ground. The grid is at a variable negative voltage with respect to the cathode. A fixed positive voltage of several thousand volts is connected to the second (accelerating) anode. The potential of the first (focusing) anode is less positive than the potential of the second anode. It can be varied to place the focal point of the electron beam on the screen of the tube. Control-grid potential is established at the proper level to allow the correct number of electrons through the gun for the desired intensity.

ELECTRON-BEAM DEFLECTION SYSTEM

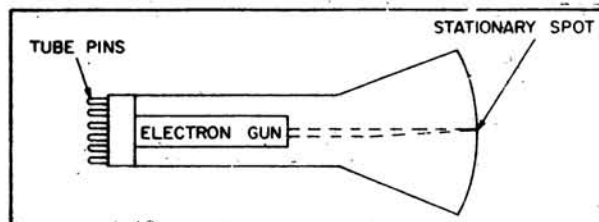
The electron beam is developed, focused, and pulled toward the screen by the electron gun. It appears on the screen of the CRT as a small, bright dot. If the beam were left in one position, the electrons would soon burn away the illuminating coating in that one area. To be of any use, the beam must move. As you have learned, an electrostatic field can bend the path of a moving electron, or an electron stream.

Assume the beam of electrons passes through an electrostatic field between two plates. Since electrons are negatively charged, they will be deflected in the direction of the electric force (from negative to positive). The electrons will follow a curved path through the field. When the electrons leave the field, they will take a straight path to the screen at the angle at which they left the field. Although the beam is still wide (the focal point is at the screen), all the electrons will be traveling toward the same spot. This is assuming, of course, that the proper voltages are existing on the anodes which produce the electrostatic field. Changing the voltages changes the focal point of the beam.

ELECTRON GUN FIELDS



CRT WITHOUT DEFLECTION



QUESTION

- Q10. Why are the electrostatic fields between electron-gun elements called lenses?

ANSWER

- A10. They are called lenses because the fields con-

Understanding Electronics

concentrate and focus the electron streams in the same manner that optical lenses bend light rays.

VERTICAL AND HORIZONTAL PLATES

If two sets of deflection plates are placed at right angles to each other inside a CRT, the electron beam can be controlled in any direction.

By varying the voltage between the two vertical-deflection plates, the spot on the face of the tube can be made to move up and down. The distance will be proportional to the change in voltage between the plates. Changing the voltage difference between the horizontal-deflection plates will cause the beam to move a given distance from one side to the other. There are directions other than up-down and left-right. The beam must be deflected in all directions.

Note the double diagram. You can see that the beam may be moved to any position on the screen simply by moving it both vertically and horizontally.

In the top diagram, position A of the beam is in the center. It can be moved to position B by going up two units and then right two units. Movement of the beam is the result of the simultaneous action of both sets of deflection plates. The electrostatic field between the vertical plates moves the electrons up an amount proportional to two units at the screen. As the beam passes between the horizontal plates, it is moved to the right an amount proportional to two units.

QUESTION

Q11. In the right figure, how many units, and in which direction, will each set of deflection plates move beam A' to B'?

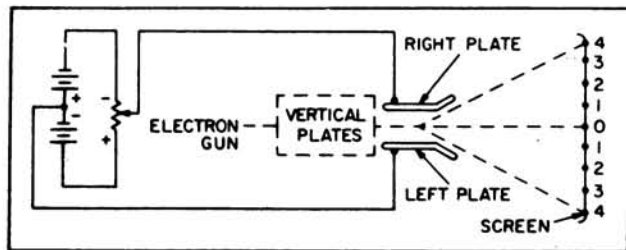
ANSWER

A11. The vertical plates will move the beam down three units. The horizontal plates will move the spot one unit to the left.

VOLTAGE CONTROL OF HORIZONTAL PLATES

Assume that the resistance of the potentiometer in the figure is spread evenly along its length. When the arm of the potentiometer is at the middle position, there is the same potential on each plate. Since there is zero potential between the plates, an electrostatic field is not produced. The beam will be at zero on the screen. If the arm is moved downward at a uniform rate, the right plate will become more positive than the left. The electron beam will move from 0 through 1, 2, 3, and 4 in equal time intervals. If the potentiometer arm moves at the same rate in the other direction, the right plate will decrease in positive potential. The beam returns to the zero position when the poten-

HORIZONTAL PLATES—TOP VIEW



tial difference between the plates again becomes zero. Moving the arm toward the other end of the resistance will cause the left plate to become more positive than the right. The direction of the electric force reverses, and the beam moves from 0 through 4'. If the movement of the potentiometer arm is at a linear (uniform) rate, the beam will move at a steady rate.

AMPLITUDE VERSUS TIME

Do you recall the statement made earlier that waveforms could be described in terms of amplitude and time? You have just seen how the movement of the beam depends on both potential (amplitude) and time.

From zero time to 1-second, the waveform in the diagram is at zero volts. In the CRT the vertical plates remain at the same potential difference while the potential difference between the horizontal plates increases 1 unit in the direction necessary to move the beam toward the right. When time is equal to 1-second, the waveform rises to +2-volts. The potential difference between the vertical plates increases enough to move the electron beam 2 units in the positive direction. From 1 to 4-seconds, the waveform remains at +2-volts and then decreases to -2-volts. As the horizontal-plate potential difference increases by 3 units, the vertical potential remains the same (+2 units) and then drops sharply 4 units. For the next 3-seconds, the waveform remains at -2-volts. In the CRT, the potential difference between the vertical plates remains unchanged as the horizontal potential increases uniformly.

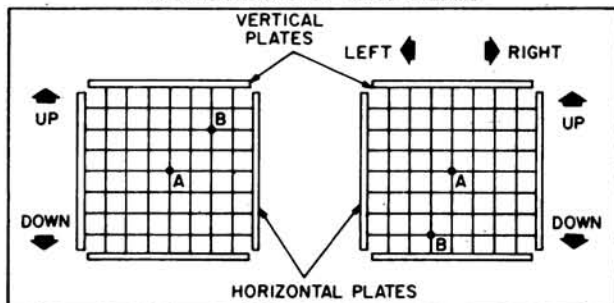
QUESTIONS

- Q12.** Waveforms can be described in terms of a _____ and t _____.
Q13. The horizontal-deflection plates are used to reproduce the a _____ t _____. (choose one)

ANSWERS

- A12.** Waveforms can be described in terms of amplitude and time.
A13. The horizontal-deflection plates are used to reproduce the time component.

DEFLECTION OF CRT BEAM



WHAT YOU HAVE LEARNED

An electron gun contains a cathode (to emit electrons), a control grid (to control the intensity of the trace on the screen), a first anode (to develop the electric lenses that focus the beam on the screen), and a second anode (to accelerate the electrons toward the screen). Deflection plates in vertical and horizontal pairs are used to position the beam on the screen. ■

